

Evaluation von phänomenbasiertem Physikunterricht

—

Seine Merkmale und Wirkungen auf
Stereotypen, Selbstkonzept und Interesse

D i s s e r t a t i o n

zur Erlangung des akademischen Grades

doctor rerum naturalium

(Dr. rer. nat.)

im Fach Physik

eingereicht an der

Mathematisch-Naturwissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin

von

Nico Westphal

Präsident der Humboldt-Universität zu Berlin

Prof. Dr. Jan-Hendrik Olbertz

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät

Prof. Dr. Elmar Kulke

Gutachter: 1. Lutz-Helmut Schön
 2. Johannes Grebe-Ellis
 3. Volkhard Nordmeier

Tag der mündlichen Prüfung: 23.07.2014

Abstract

Phänomenbasierter Physikunterricht (PbPU) als alternative Vorgehensweise bei der Vermittlung von Physik wird seit 20 Jahren in der fachdidaktischen Forschung verstärkt beachtet und hat sich zunehmend an den Schulen etabliert. Er lässt sich anhand seiner Merkmalaspekte Subjektivität, Mediation, Affektivität, Modellfreiheit/Modellkompetenz und Exploration charakterisieren und abgrenzen. Die Merkmale PbPUs sind mit einer konstruktivistischen Lerntheorie vereinbar und liefern die Voraussetzungen für erfolgreiches Lernen gemäß dem INVO-Modell. Bisher existieren jedoch keine empirischen Belege für den Einfluss PbPUs auf die Interessen- und Leistungsentwicklung oder anderer motivationaler Konstrukte wie Stereotypen und das fachbezogene Selbstkonzept. Für den wertenden Vergleich mit anderen Konzeptionen von Physikunterricht ist es jedoch wichtig, PbPU hinsichtlich seiner Wirkung auf das Lernen von Physik einschätzen zu können. Mit der vorliegenden Arbeit soll diesem Desiderat begegnet werden.

In einer quasiexperimentellen Feldstudie im Versuchs-/Kontrollgruppen-design ($N = 300$) wurden die Vor- und Nachtestergebnisse der Versuchsbedingungen verglichen. Das Treatment in der Versuchsgruppe war eine achtwöchige Unterrichtssequenz der phänomenbasierten Anfangsoptik (Klassenstufe 7). Als abhängige Konstrukte wurden das Image von Physik, das fachbezogene Selbstkonzept, Interesse und Fachwissen untersucht. Image und Selbstkonzept wurden im Kontext der Balanced Identity Theory operationalisiert und mithilfe impliziter Assoziationstests (IATs) erhoben.

Die varianzanalytischen Verfahren zeigten keine Effekte, die auf einen von der Testzeit abhängigen Unterschied zwischen der Versuchs- und Kontrollgruppe hindeuten. Die Ausprägungen der impliziten Assoziationen und des Interesses entwickelten sich in beiden Gruppen identisch, mit einer für das Lernen von Physik ungünstigen Tendenz. Die Ergebnisse dieser Studie liefern somit keine Belege für die Wirksamkeit PbPUs auf die untersuchten Konstrukte.

Abstract

Since 20 years are "phenomenon-based science classes" (PbPU) as alternative method for imparting physics increasingly observed by subject-didactic research and they have established for use in schools. PbPU can be characterized and defined by the specific aspects subjectivity, mediation, affectivity, model-competency and exploration. Defining characters of PbPU are compatible with constructivist theory of learning and provide conditions for effective learning in accordance to INVO-model. Until today there are no empirical demonstrations to state the effect of development of PbPU in interests and performance nor other motivational constructs as stereotypes and physics self-concept. It is important for evaluative comparison with different concepts of science classes to be able to value PbPU concerning effect on learning of physics. The work in hand shall respond to this desideratum.

Pre- and posttest data of two experimental conditions were compared in nonrandomized field study with control group design ($N = 300$). Treatment of test group was an eight weeks' teaching sequence of phenomenon-based optics (7th grade). The dependent constructs were image of physics, physics self-concept, interests and knowledge in optics. Image and self-concept were operationalized within theoretical framework of Balanced Identity and measured with implicit association tests (IATs).

ANOVAs and ANCOVAs revealed no effects of interaction between time and experimental group. Implicit associations and interests developed in each group identically with negative tendency for learning of physics. Therefore, results of this study provide no evidence for the impact of PBPU on investigated constructs.

Vorwort

Die vorliegende Arbeit beinhaltet die Darstellung meiner mehrjährigen Auseinandersetzung mit der Evaluation phänomenbasierten Physikunterrichts. Das Gelingen eines solchen Projektes ist von vielen Faktoren, vor allem aber der Unterstützung und Expertise weiterer Personen abhängig. Ich hatte das Glück, beides von vielen Wegbegleitern in Form von Rat und Tat genießen zu dürfen. Dem Umstand, dass mir die Arbeit der letzten Jahre ohne diese Personen unschätzbar schwerer gefallen wäre, soll mit diesem Vorwort angemessene Beachtung geschenkt werden.

Mein erster Dank geht an Prof. Dr. Lutz-Helmut Schön. Ich verdanke ihm als Lehrer und Betreuer eine besondere Sicht auf das Lehren von Physik, die ich nicht missen möchte. Die Balance zwischen der von ihm geförderten eigenständigen und eigenverantwortlichen Arbeit auf der einen sowie spontanen Hilfestellung zu jeder Zeit auf der anderen Seite, habe ich als sehr angenehm empfunden. Gleichzeitig verbreitete er die nötige Portion Optimismus, wodurch die gemeinsamen Gespräche für mich stets beruhigend endeten.

Des Weiteren möchte ich mich bei Prof. Dr. Johannes Grebe-Ellis bedanken. Das Spektrum des mir von ihm vermittelten fachdidaktischen Wissens geht über die phänomenologische Methode der Naturerschließung weit hinaus, womit ein Grundstein für die eigenen Forschungstätigkeiten auf dem Gebiet gelegt wurde. Er war es nie müde, detailliert zu beraten sowie Rede und Antwort zu Fragen in schwer zugänglichen Themengebieten zu geben.

Schließlich bedanke ich mich außerdem bei Prof. Dr. Burkhard Priemer, der mir die Möglichkeit bot, mit genügend Freiraum für die Forschung, mein Promotionsprojekt in seiner Arbeitsgruppe fortzuführen.

Die Finanzierung des Projekts, genauso wie inhaltliche Unterstützung in Form eines Begleitprogramms, verdanke ich dem Bundesministerium für Bildung und Forschung.

Eine an Schulen durchgeführte Studie ist ohne die Unterstützung von Lehrkräften nicht möglich. Daher möchte mich zunächst bei Steffen Harke bedanken, der mir sehr bei der Suche nach Studienteilnehmern geholfen hat und mir bei vielen Dingen die praktische Durchführung an seiner Schule erleichterte. Außerdem geht mein Dank an Kristine Bittner, Frank Breuer und Rainer Großmann. Sie waren bereit, ihre Klassen in meine Obhut zu geben und erwiesen sich damit durchaus als experimentierfreudig und offen für Unbekanntes. Das Gleiche gilt für alle teilnehmenden Lehrer, die hier nur aus Gründen der Anonymität nicht namentlich genannt werden. Ich danke ihnen und ihren Klassen, die ohne Klage die durchgeführten Erhebungen über sich ergehen ließen.

Von Beginn an haben mich auf meinem Weg Dr. Franz Boczianowski und Marc Müller begleitet. Mit Franz als Bürokollege war die Anwesenheit am Schreibtisch durch humorvolle Gespräche über das Wesentliche stets angenehm. Marc lernte ich als scharfsinnigen Kritiker zu schätzen, der gern bereit war, an den Fundamenten meiner Argumente zu rütteln, um diese auf ihre Stabilität zu prüfen. Beiden gilt besonderer Dank für die langjährige gute Zusammenarbeit in der Lehre und für den Austausch über die wissenschaftlichen Arbeit. Darüber hinaus bedanke ich mich bei allen ehemaligen und aktuellen Kolleginnen und Kollegen der AG Didaktik der Physik für ihre Zusammenarbeit und Unterstützung in den unterschiedlichsten Bereichen.

Schließlich bleibt der Dank an meine Familie, die den nötigen Rückhalt bot und ohne ständiges Nachfragen auskam, wann man den endlich mit einer richtigen Arbeit beginne. Insbesondere danke ich Anett Grasnick für zahlreiche Probevorträge auf der Couch, Ausflüge in die Spiegelwelt und Beistand in schwierigen Stunden.

Berlin, 17. Juli 2014

Nico Westphal

Inhaltsverzeichnis

Vorwort	vii
1 Einleitung	1
 I Theoretischer Teil	
2 Phänomenbasierter Physikunterricht	7
2.1 Zur Bezeichnung <i>phänomenbasiert</i>	7
2.2 Merkmale phänomenbasierten Physikunterrichts	9
2.3 Lerntheoretische Verortung	24
2.4 Ziele phänomenbasierten Physikunterrichts	30
2.5 Phänomenbasierte Anfangsoptik	31
2.6 Bedeutung für die vorliegende Arbeit	39
3 Assoziationen zum Schulfach Physik	41
3.1 Interesse an Physik	41
3.2 Image von Physik	46
3.3 Assoziationen zu Mathematik und Naturwissenschaften	52
3.4 Bedeutung für die vorliegende Arbeit	54
4 Messung von impliziten Assoziationen	57
4.1 Implicit Association Test	58
4.1.1 Ablauf eines IATs	58
4.1.2 Berechnung des Assoziationsmaßes (<i>D</i> -Wert)	63
4.1.3 Psychometrische Eigenschaften	65

4.1.4	Bisheriger Einsatz im Kontext <i>Naturwissenschaften</i> und <i>Schule</i>	72
4.2	Single Target IAT	73
4.2.1	Ablauf eines ST-IATs und Berechnung des <i>D</i> -Werts	74
4.2.2	Psychometrische Eigenschaften	76
4.3	Bedeutung für die vorliegende Arbeit	78
5	Balanced Identity Theory	79
5.1	Definitionen, Prinzipien und empirische Überprüfung der BIT	80
5.2	Balanced Identity und das Image von Physik	84
5.3	Bedeutung für die vorliegende Arbeit	87
6	Forschungsfragen und Hypothesen	89
6.1	Entwicklung der Forschungsfragen	89
6.2	Hypothesen	91
 II Empirischer Teil		
7	Untersuchungsmethode	97
7.1	Wahl des Forschungsdesigns und der Analysemethode	97
7.2	Wahl der Instrumente	101
7.3	Statistische Hypothesen	105
8	Treatment	107
9	Pilotstudien	123
9.1	Erste Pilotierung	123
9.2	Zweite Pilotierung	125
10	Hauptstudie: Evaluation der phänomenbasierten Anfangsoptik	129
10.1	Beschreibung der Stichprobe	129
10.2	Datenerhebung	130
10.3	Subjektive Eindrücke von den Erhebungen	131
10.4	Datenaufbereitung, Testgüte und Skalierung	132
11	Ergebnisse	137
11.1	Überprüfung der Vorannahmen	137
11.2	Implizite Assoziationen	140
11.3	Explizite Assoziationen	143
11.4	Interesse	145
11.5	Fachwissen	149
11.6	Weiterführende Untersuchungen	150

III Schluss

12 Diskussion und Ausblick	157
12.1 Assoziationen zu Physik	157
12.2 Interesse und Fachwissen	160
12.3 Allgemeine Interpretation	162
12.4 Ausblick	165
13 Zusammenfassung	167
Anhang	170
A Testinstrumente	171
B Unterrichtsmaterialien	197
C Grafiken, Tabellen und Formeln	211
Abbildungsverzeichnis	231
Tabellenverzeichnis	233
Literatur	235

KAPITEL 1

Einleitung

Im Laufe der letzten 30 Jahre hat sich in der deutschsprachigen Physikdidaktik eine Forschungsrichtung herausgebildet, die im Allgemeinen als *phänomenologisch*, *phänomenorientiert* oder auch *erscheinungsorientiert* bezeichnet wird. Die dazugehörigen und maßgeblich durch die Pädagogen Martin Wagenscheins sowie der Waldorfschulen beeinflussten Konzeptionen von Physikunterricht reichen von Curricula des Anfangsunterrichts bis zu Vorschlägen der modernen Quantenphysik. Sie haben alle die Gemeinsamkeit, dass die Phänomene Ausgangspunkt, Fundament und auch Endpunkt der Auseinandersetzung mit der Natur sind. Trotz einiger Unterschiede der Arbeiten, z. B. bei erkenntnistheoretischen Grundhaltungen, gibt es einen Konsens bezüglich vieler Implikationen für den Physikunterricht. Dieser Konsens wird hier als *phänomenbasierter Physikunterricht (PbPU)* bezeichnet.

Die Profilierung des phänomenbasierten Physikunterrichts geht aufgrund verstärkter Forschungsaktivitäten mittlerweile über den deutschsprachigen Raum hinaus. Mit dem Ziel, seine Tragfähigkeit auch für das Niveau der Sekundarstufe II nachzuweisen, entstanden in den letzten 10 Jahren zahlreiche phänomenologische Erschließungen neuer Themenbereiche. In diesem Zuge hat sich die phänomenbasierte Anfangsoptik als ein Beispiel für PbPU im schulischen Alltag stetig weiter etabliert. Ihre Inhalte haben Eingang in Rahmenlehrpläne gefunden und sind in Schulbüchern verschiedener Bundesländer berücksichtigt. Aus der Literatur und von vielen Rückmeldungen aus der Praxis sind viele Vorzüge PbPUs gegenüber konventionellen Physikunterrichts bekannt. Teilweise scheinen sie evident zu sein und teilweise lassen sie sich auch aus lerntheoretischer Sicht begründen (siehe Abschnitt 2.3). Bisher gibt es jedoch kaum eine empirische Fundierung dafür, dass PbPU sich auf die Interessen und Leistungsentwicklung der Lernenden auswirkt. Dies ist vor allem deshalb ein Mangel, da die Interessenförderung ein explizit formuliertes Ziel der phänomenologisch orientierten Physikdidaktik ist.

Das Interesse an Physik ist ein sehr detailliert erforschtes Konstrukt. Zahlreiche Studien, wie die IPN-Interessenstudie, PISA oder der IQB-Ländervergleich lassen differenzierte Aussagen über den Verlauf des Interesses, Geschlechtsunterschiede und den Vergleich mit anderen Fächern zu. Sie zeigen im Wesentlichen das einheitliche Bild, dass 1. das Interesse an Physik niedriger ist, als bei anderen Fächern, 2. bei Mädchen sehr viel geringer ist, als bei Jungen und 3. im Verlauf der Schulzeit immer weiter sinkt. Mit dem Interesse geht eine dementsprechende Ausprägung des physikbezogenen Selbstkonzepts und der Beliebtheit bei den Schülerinnen und Schülern einher. Genauso existieren Assoziationen zu Physik, die dem Fach ein Image als *schwierig* und *männlich* zuschreiben. Diese Assoziationen behindern die Entstehung eines fachbezogenen Selbstkonzepts und damit die Förderung des Interesses.

Das Motiv der vorliegenden Arbeit besteht darin, einen empirischen Beleg für die Wirkung phänomenbasierten Physikunterrichts auf das Lernen von Physik zu geben. Dafür sollen die Assoziationen zum Fach als Folge des Unterrichts in Betracht gezogen werden. Anhand ihrer kann der Einfluss auf das Image von Physik und das fachbezogene Selbstkonzept ermittelt werden. Für die Messung der Assoziationen kommen unter anderem implizite Verfahren zum Einsatz, die es ermöglichen die gedanklichen Verknüpfungen auf unbewusster Ebene zu erfassen. Schließlich wird außerdem der Einfluss auf das Interesse bestimmt, um zu kontrollieren, ob die angestrebte Förderung durch PbPU gewährleistet wird. Als viertes Kriterium für die Wirkung auf das Lernen von Physik wird letztendlich auch das Fachwissen beobachtet.

In Kapitel 2 wird zunächst phänomenbasierter Physikunterricht als Untersuchungsgegenstand dieser Studie vorgestellt. Die in der bestehenden Literatur identifizierbaren Merkmale PbPUs werden geordnet nach fünf Aspekten (*Subjektivität, Mediation, Affektivität, Modellfreiheit/Modellkompetenz* und *Exploration*), anhand derer sich der Unterricht charakterisieren und abgrenzen lässt. Diese explorative Ermittlung der Merkmale wird gestützt durch eine vorab stattgefundene Befragung, in der Experten Aussagen über den Unterricht bewertet haben. Anschließend wird PbPU aus Sicht einer konstruktivistischen Auffassung vom Lernen bewertet und in einem darauf aufbauenden Modell über individuelle Voraussetzungen erfolgreichen Lernens verortet. Mit Blick auf die geplante Studie wird der Untersuchungsgegenstand schließlich auf die phänomenbasierte Anfangsoptik eingeeengt.

Kapitel 3 legt die Forschungslage zu den Konstrukten dar, die als abhängige Variablen untersucht werden. Zunächst wird das Konstrukt *Interesse* definiert. Anschließend werden grundlegende Befunde der bisherigen Studien aufgeführt und Verbindung zum Image von Physik und zum fachbezogenen Selbstkonzept gebracht. Außerdem werden der Zusammenhang der drei Konstrukte mit den Assoziationen zu Physik erörtert und diesbezügliche Befunde vorgestellt.

Da es sich bei den eingesetzten Verfahren zu Messung der impliziten Assoziationen um in der fachdidaktischen Forschung unübliche Tests handelt, werden in Kapitel 4 der Implicit Association Test (IAT) und der Single Target IAT vorgestellt. Die Rechtfertigung

tigung für den Einsatz des „neuen“ Testverfahrens geschieht durch eine umfangreiche Diskussion der Validität des Tests anhand diesbezüglicher Studien. Ferner wird aufgezeigt, inwiefern der Test bereits erfolgreich im Kontext *Naturwissenschaften* und *Schule* eingesetzt wurde.

Kapitel 5 rundet mit der Einführung der *Balanced Identity Theory (BIT)* die zwei vorherigen Kapitel ab. Im Rahmen der Theorie werden die verwendeten Konstrukte *Image* und *Selbstkonzept* mit den IAT-Ergebnissen in Zusammenhang gebracht. Außerdem erlaubt sie Vorhersagen über die Wechselwirkungen zwischen den Assoziationen. Beides zusammen ermöglicht eine theoriegeleitete Beschreibung der in Kapitel 3 erörterten Zusammenhänge mithilfe impliziter Assoziationen.

In Kapitel 6 werden die bisher dargelegten theoretischen Aussagen zusammen mit dem bereits oben angesprochenen Forschungsdesiderat verwoben und daraus vier Forschungsfragen generiert. Aus der lerntheoretischen Verortung PbPUs und der BIT heraus lassen sich schließlich theoriegeleitet die Hypothesen der vorliegenden Studie entwickeln.

Die Forschungsfragen implizieren das Forschungsdesign und die Analysemethoden, die durch eine knappe Diskussion der Alternativen begründet werden. Außerdem werden die verwendeten Instrumente und das in der Versuchsgruppe durchgeführte Treatment vorgestellt (Kapitel 7 und 8). Aus der Wahl des Forschungsdesigns, der Analysemethoden und der Instrumente ergeben sich schließlich die statistischen Hypothesen (Abschnitt 7.3).

In Kapitel 9 werden die zwei durchgeführten Pilotstudien mit ihren Ergebnissen dargestellt. Sie hatten maßgeblichen Einfluss auf die in der Hauptstudie verwendeten Instrumente und sind daher gesondert aufgeführt. In Kapitel 10 erfolgt die Beschreibung der Hauptstudie. Sie beinhaltet die deskriptiven Statistiken der Stichprobe, den Ablauf der Erhebungen, die Verfahren der Skalierung sowie Berechnung der in den weiteren Analysen verwendeten Kennwerte. In Abschnitt 10.3 werden außerdem subjektive Eindrücke geschildert, die der Autor der vorliegenden Arbeit als Versuchsleiter bei den Erhebungen gewonnen hat. Sie sollen es dem Leser ermöglichen, die Ergebnisse auch unter Berücksichtigung dieser Aspekte zu bewerten.

Im Anschluss an die Darstellung der Ergebnisse, die sich auf die Überprüfung der Hypothesen beziehen, folgt ein kurzer explorativer Abschnitt mit weiterführenden Untersuchungen (Kapitel 11 und Abschnitt 11.6). Es folgt die Diskussion der Ergebnisse mit der Beantwortung der Forschungsfragen (Abschnitte 12.1 und 12.2) und eine allgemeine Diskussion in Bezug auf die grundlegende Motivation der Arbeit (Abschnitt 12.3). Den Abschluss bilden ein kurzer Ausblick auf weitere Forschungen, die das Thema der Arbeit betreffen (Abschnitt 12.4) sowie eine Zusammenfassung der gesamten Arbeit (Kapitel 13).

Teil I

Theoretischer Teil

Phänomenbasierter Physikunterricht

*Was nicht auf den Phänomenen steht, wird
nicht verstanden und deshalb schnell
vergessen.*

Martin Wagenschein

Das Zitat von Martin Wagenschein eignet sich in zweierlei Hinsicht, um das Anliegen dieses Kapitels zu skizzieren. Zum einen deutet es an, was hier unter *phänomenbasiert* verstanden werden soll und zum anderen gibt es einen Hinweis auf das Motiv und die Zielsetzung eines derartigen Physikunterrichts. An dieser Reihenfolge orientierend, wird zunächst der *phänomenbasierte Physikunterricht* definiert und charakterisiert, indem Beziehungen zu bestehenden Unterrichtskonzeptionen hergestellt werden und außerdem der Versuch unternommen wird, kennzeichnende Merkmale zu nennen. Darauf fußend soll eine Verortung in bestehende Lehr-/Lerntheorien vorgenommen werden. Aus dieser Verortung heraus und anhand der existierenden Vorarbeiten werden die Beweggründe und Ziele der phänomenbasierten Vorgehensweise im Physikunterricht gesammelt. Aus ihnen soll es letztendlich möglich sein, prüfbare Hypothesen zum Gelingen von PbPU abzuleiten. Ferner wird das Sichtfeld auf die detailliert entwickelte und etablierte *phänomenbasierte Anfangsoptik* eingeengt, die den Untersuchungsgegenstand der vorliegenden Arbeit darstellt. An ihr werden die Merkmale und Ziele des phänomenbasierten Vorgehens inhaltlich konkretisiert.

2.1 Zur Bezeichnung *phänomenbasiert*

In der naturwissenschaftsdidaktischen Forschung haben sich die Begriffe *phänomenologisch* (z. B. Mackensen, 1992; Wagenschein, 2009d; Weber, 2003; Grebe-Ellis, 2005),

phänomenorientiert (z. B. Krüger, 2011a) und *erscheinungsorientiert* (Theilmann, 2006b) etabliert, deren begriffliche Ähnlichkeit auch inhaltliche Gemeinsamkeiten vermuten lassen. Solche Gemeinsamkeiten lassen sich auch in Form eines Konsens über eine didaktische Grundhaltung finden. Jedoch gibt es auch Unterschiede, die sich aus verschiedenen philosophischen Grundhaltungen und aus unterschiedlichen erkenntnis- und lerntheoretischen Verortungen ergeben. Deutlich wird dies schon bei der vielfältigen Nutzung des Begriffes *phänomenologisch* in den Naturwissenschaften, in der Philosophie und in der Physikdidaktik, was bereits durch Grebe-Ellis (2005) dargelegt wird. Zusätzlich geben Østergaard, Dahlin und Hugo (2008) eine Übersicht über die Anwendungen von *phenomenology in science education* als Erkenntnistheorie oder als qualitativen Forschungsansatz in den Sozialwissenschaften, einschließlich der Fachdidaktiken. Die noch zu entwickelnde Fragestellung der vorliegenden Arbeit bezieht sich jedoch allein auf die Konzeption von Physikunterricht. Es soll keine Frage in der Art beantwortet werden, ob *Phänomenologie* als Form der Naturerschließung oder als Forschungsmethode in der Physikdidaktik möglich, sinnvoll oder gar nötig ist.

In Bezug auf den Physikunterricht lassen sich im Wesentlichen zwei Ursprünge identifizieren, aus denen die oben genannten didaktischen Konzeptionen hervorgehen. Das ist zum einen das umfangreiche Werk von Martin Wagenschein und zum anderen die Waldorfpädagogik. Wagenscheins wissenschaftliche Arbeit umfasst einerseits zahlreiche episodische Erzählungen, Vorträge sowie beispielhafte Unterrichtslehrgänge (z. B. Wagenschein, 2009c) und andererseits Veröffentlichungen mit dem Schwerpunkt, die von ihm vertretende Art des Physiklehrens (*genetisch-sokratisch-exemplarisch*) darzulegen (z. B. Wagenschein, 1976, 2010)¹. Seine Arbeiten finden bis heute in der Schulpraxis und der fachdidaktischen Forschung große Beachtung und wurden auch kontrovers diskutiert (z. B. Engelbrecht, 2003; Muckenfuß, 1995). Für beispielhafte Darstellungen der Waldorfpädagogik seien hier Mackensen (1992), Mackensen und Ohlendorf (1998) und die für den Bereich der Optik grundlegenden Schriften von Maier (2003, 2004) genannt. Auch für die Waldorfpädagogik existieren zahlreiche wissenschaftliche Auseinandersetzungen, auf die an dieser Stelle nur exemplarisch verwiesen werden kann (z. B. Barz & Randoll, 2007; Jelinek & Sun, 2003; Paschen, 2010). Die Didaktik Wagenscheins und die Waldorfpädagogik haben unterschiedliche Ausgangspunkte für ihre Konzeptionen des Physikunterrichts und gehen „durchaus deutlich verschiedene Wege“ (Buck und Mackensen, 2006, S. 4). Jedoch gibt es auch viele, bereits oben angesprochene Verwandtschaften und Gemeinsamkeiten, die sich umso mehr im Resultat der didaktischen Bemühungen widerspiegeln – den konkreten Vorschlägen für die Gestaltung von Physikunterricht.² Dieser physikdidaktische Konsens soll hier als *phänomenbasierter Phy-*

¹Einen umfassenden Überblick über das Lebenswerk von Martin Wagenschein findet man auf den Seiten des Internetauftritts vom Martin Wagenschein-Archiv (Kohl, 2013)

²Auf eine ausführliche Darstellung der Beziehung zwischen Waldorf- und Wagenscheinpädagogik und deren Wurzeln wird im Rahmen dieser Arbeit verzichtet, da sie für die noch zu entwickelnde Fragestellung nicht relevant ist. Für einen Überblick zu der Thematik sei auf Buck und Mackensen (2006) verwiesen.

sikunterricht (PbPU) bezeichnet werden. Die Wahl der neuen³ Bezeichnung erleichtert das Erinnern an diesen Fokus. Als PbPU kann sich demnach die Schnittmenge der oben genannten Konzeptionen vorgestellt werden, die sich auf Unterrichtsgestaltung, Unterrichtsinhalte sowie Art und Weise der Wissenskonstruktion seitens der Schülerinnen und Schüler bezieht.

Ein weiterer Grund für die Bezeichnung *phänomenbasiert* liegt darin, dass sich PbPU nicht an Phänomenen oder Erscheinungen orientiert, indem er sich nach ihnen ausrichtet oder sie sogar als Ziel physikalischen Handelns begreift. Im Extremfall könnten so die Phänomene am Ende nur als Beispiel für die im Unterricht gewonnene Erkenntnis dienen. Vielmehr dienen die Phänomene dem PbPU als *Basis*, „um auf diesem *Fundament* ihr [das der Physik] begriffliches Gebäude zu errichten“ (Wagenschein, 1989, S. 109), so wie auch die Säulen einer Kathedrale auf einer *Basis* stehen. Genauso ist das *Basislager* der Ausgangs- und letztendlich auch wieder Endpunkt einer Expedition, als welche man sich bildlich den Physikunterricht vorstellen kann. Eigentliches Ziel dieser Expedition (der Gipfel oder ein anderer entlegener Ort) ist jedoch die „Klärung“ des Phänomens mit physikalischen Mitteln. In der Mathematik stellt die *Basis* eines Vektorraums ein minimales Erzeugendensystem von Vektoren dar, aus dem sich der gesamte Vektorraum „aufspannen“ lässt. Stellt man sich den Physikunterricht als einen Vektorraum vor, so wären im PbPU die Natur- und Laborphänomene die Basis. Jedes Element (Vektor) in diesem Raum, z. B. eine physikalische Größe oder ein Gesetz, ist eine Linearkombination aus den Basisvektoren, lässt sich also aus den Phänomenen „erzeugen“. Dieser Vergleich hinkt natürlich insofern, als dass man sich physikalische Theorien, wie mathematische Theorien, auch unabhängig von einer realen Wirklichkeit, also unabhängig von Phänomenen vorstellen könnte. Dies geschieht im phänomenbasierten Physikunterricht aber gerade nicht oder nur sehr selten. Hier muss es für die Schülerinnen und Schüler zu jeder Zeit möglich sein, die Elemente des Physikunterrichts auf Phänomene der realen (Alltags-)Welt zurückzuführen, also zum Basislager zurückzukehren.

Diese zum Teil sehr metaphorische Begründung dient der Rechtfertigung für die Einführung einer neuen Bezeichnung, die anstelle der bereits etablierten Begriffe verwendet werden soll. Zum einen soll sie es erleichtern, in einem komplexen Forschungsfeld den Überblick darüber zu behalten, worin der Untersuchungsgegenstand der vorliegenden Arbeit besteht. Zum anderen ist sie m. E. geeigneter als bisher gewählte Formulierungen, da sie eine grundlegende Charakteristik des Unterrichts im Sinne des Zitats von Wagenschein (S. 2) einschließt.

2.2 Merkmale phänomenbasierten Physikunterrichts

Die Merkmale phänomenbasierten Physikunterrichts können (i) mithilfe der Ergebnisse einer eigenen zweistufigen Expertenbefragung (vgl. Westphal, Schön & Grebe-Ellis,

³Die Bezeichnung *phänomenbasiert* ist nur in diesem Kontext neu und wird bereits von Aufschnaiter (2005) in einem anderen Zusammenhang zu Kategorisierung von Schülerkonzepten verwendet.

2011) und (ii) anhand vieler phänomenbasierter Vorarbeiten gesammelt werden. Die Expertenbefragung (i) bestand in der ersten Stufe darin, einem ausgewählten Kreis von Vertretern⁴ ($N = 10$) ein Beispiel einer typisch phänomenbasierten Unterrichtseinheit in Form eines offenen Protokolls skizzieren zu lassen. So sollten explorativ Unterrichtsformen, Inhalte, Methoden, Schwerpunktsetzungen und Ziele identifiziert werden. In der zweiten Stufe wurden die so ermittelten Eigenschaften in Form von Aussagen einem erweiterten Kreis von Experten ($N = 25$) vorgelegt. Zu den *Vertretern* kamen *Lehrer* und *Externe* hinzu. Die Gruppe der Lehrer stammt aus dem Umfeld der AG Didaktik der Physik der HU-Berlin. Sie kennt PbPU durch Teilnahme an den Projekten SINUS⁵ und PiKo⁶, durch Lehrerfortbildungen oder durch Mitarbeit im UniLab Adlershof. Wesentliche Leitideen von PbPU und einige Unterrichtssequenzen werden von ihnen bereits seit vielen Jahren in der Schule berücksichtigt. Als *Externe* wird eine Gruppe von Fachdidaktikern bezeichnet, von denen angenommen wird, dass sie PbPU durch wissenschaftlichen Austausch sehr gut kennen und charakterisieren können. Im Gegensatz zu den Vertretern befürworten sie PbPU nicht unbedingt in allen Facetten bzw. lehnen sie ihn sogar teilweise ab. Der erweiterte Kreis von Experten sollte die Aussagen über die Eigenschaften geschlossen in Form von Likert-Skalen bewerten. Um auf einer breiteren Basis an möglichen Eigenschaften zu stehen, wurden die Items der geschlossenen Befragung zusätzlich zu den Ergebnissen der ersten Stufe ergänzt um Aussagen, die aus den phänomenbasierten Vorarbeiten (ii) generiert werden konnten.

Unter den veröffentlichten Vorarbeiten lassen sich nur wenige Beiträge finden, in denen phänomenbasierter Unterricht direkt charakterisiert wird (z. B. Theilmann, 2008; Krüger, 2011b; Grebe-Ellis, 2006a). Der größte Teil thematisiert mehr oder weniger konkrete Lehrgänge, aus denen die Merkmale nur implizit abgeleitet werden können (z. B. Buck & Mackensen, 2006; Mackensen, 1992; Schön, 1985, 1994; Wagenschein, 2009c). Im Folgenden wird versucht, diese gesammelten Merkmale systematisch geordnet darzustellen und, sofern möglich, ihnen anhand der Bewertungen der Expertenbefragung Gewicht zu verleihen.

Phänomenbasierter Physikunterricht lässt sich unter verschiedenen Gesichtspunkten (Aspekte) charakterisieren. Diese sind m. E. auf fünf geeignete Aspekte reduzierbar: *Subjektivität*, *Mediation*, *Affektivität*, *Modellfreiheit/Modellkompetenz* und *Exploration* (siehe Abb. 2.1). Diese fünf Aspekte sind nicht im strengen Sinne trennscharf voneinander zu verstehen, d. h. die Charakterisierung von PbPU unter einem bestimmten Aspekt ist nicht immer disjunkt zu den anderen Aspekten. So lassen sich insbesondere bei den Aspekten *Modellfreiheit/Modellkompetenz* und *Exploration*, aber auch bei *Subjektivität* und *Mediation* Merkmale finden, die sowohl im Sinne des einen als auch des anderen interpretierbar wären.

⁴Als *Vertreter* werden hier solche Experten bezeichnet, die phänomenbasierten PbPU selbst entwickelt, publiziert und vertreten haben.

⁵Steigerung der Effizienz des mathematisch-naturwissenschaftlichen Unterrichts

⁶Physik im Kontext

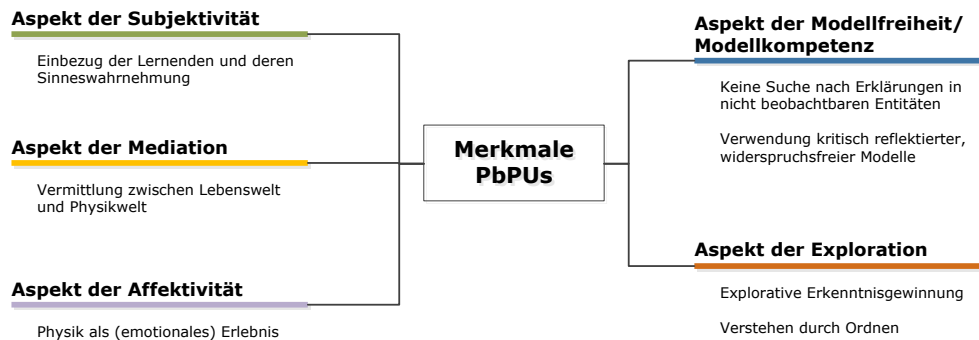


Abb. 2.1 Merkmalsaspekte von phänomenbasiertem Physikunterricht

Aspekt der Subjektivität

Der Aspekt der Subjektivität beinhaltet diejenigen Eigenschaften, die mit der Perspektive der Schülerinnen und Schüler, deren Wahrnehmung, deren Sinne, deren Erfahrungen und den Erkenntnissen, die sie daraus gewinnen, zusammenhängen. Es ist ein zentrales Merkmal von PbPU, als Ausgangspunkt jeder physikalischer Betrachtung Alltagserfahrungen oder unmittelbare Beobachtung der Schülerinnen und Schüler zu wählen, um aus ihnen heraus die Grundbegriffe der Physik zu entwickeln (Theilmann, 2006a; Weber & Schön, 2000; Schön & Weber, 2001). Als Konsequenz folgt daraus, dass der Wahrnehmung mehr Bedeutung beigemessen wird. Dieser Einbezug der sinnlichen Wahrnehmung bedeutet nicht, dass den Sinnen ausschließlich vertraut wird. Es bedeutet, dass die physikalischen Gesetzmäßigkeiten nicht unabhängig von der sinnlichen Erfahrbarkeit thematisiert werden, was auch als eindeutiges Ergebnis der Expertenbefragung festgehalten werden kann (Westphal et al., 2011). Die menschliche Wahrnehmung wird nicht von vornherein als Sinnestäuschung abgewertet, indem sie mit „objektiven“ physikalischen Messverfahren verglichen wird, um anschließend deren Notwendigkeit mit augenscheinlichen Unterschieden zu begründen. Fedra und Schön (1989) zeigen für den menschlichen Temperatursinn, dass ein solcher Vergleich mitunter nicht angemessen ist, wenn er die spezifische Funktion der Sinnesorgane übersieht. Dies bedeutet umgekehrt nicht, dass auf physikalische Messgeräte verzichtet werden kann. Denn die Bedeutung der Messgeräte liegt laut Fedra und Schön (1989) darin, dass sie sich trotz deren häufigen Ursprungs in einer Sinneswahrnehmung von diesen lösen können. Sie sind damit in der Lage, physikalische Größen zu messen (z. B. Temperatur), die in der Wahrnehmung nicht messbar sind und somit für das wahrnehmende Subjekt eigentlich nicht existieren. Für eine gedanklich konstruierte, also theoretische Größe wie der Temperatur ist demnach auch ein Messgerät nötig, das nach dieser Theorie entworfen ist. Nach Ansicht der Autoren muss jedoch im Physikunterricht deutlich werden, worin die Ursachen der möglichen scheinbaren Widersprüche zwischen Sinneswahrnehmung und „objektiver“ Messung liegen.

Im phänomenbasierten Vorgehen werden die physikalischen Grundbegriffe jedoch gerade aus der Wahrnehmung heraus entwickelt, mit dem Anspruch, auf dieser Basis eine möglichst weitgehende Naturbeschreibung im Sinne eines allgemeinbildenden Physikunterrichts zu erreichen. Das dahinterliegende Programm *Von der Wahrnehmung zur Physik* wird prägnant von Schön (1994) beschrieben: „Dabei geht es uns darum, mit den Schülern die sinnlichen Wahrnehmungen der Wirklichkeit zu untersuchen und begrifflich aufzuarbeiten und so schrittweise das Gedankengebäude der Physik zu errichten“ (S. 2). Die Realisierbarkeit dieses Ziels wurde bereits in vielen Bereichen der Schulphysik gezeigt. In der Optik können u. a. die Themen *Photometrie*, *Schatten*, *Abbildungen*, *Spiegel*, *Brechung/Hebung* und *Polarisation* vom Sehen her so entwickelt werden, dass das „Gedankengebäude der Physik“ mithilfe von Ansichten im Sinne einer „Optik der Bilder“ (Maier, 2003) entsteht (Grebe-Ellis, 2005; Mackensen, 1992; Weber & Schön, 2000). Die Mechanik ermöglicht Zugänge über den Bewegungssinn und wahrgenommenen Zug und Druck, z. B. am Hebel (Bockemühl, 1995; Buck & Mackensen, 2006; Theilmann, 2006b). In der Akustik und Thermodynamik gibt es zahlreiche Beispiele, die vom Hören bzw. Temperatursinn ausgehen und selbst die Elektrizitätslehre erlaubt erste Erfahrungen mithilfe des Geschmackssinns (Mackensen, 1992). Trotz der klaren Tendenz, der sinnlichen Wahrnehmung mehr Gewicht zu verleihen, wird laut der Expertenbefragung im PbPU dennoch versucht, die Beobachtungen auch objektiv zu analysieren bzw. die subjektive Wahrnehmung durch objektive Untersuchungen zu bestätigen.

Nach Martin Wagenschein hat man sich in der Physik der Akustik dazu entschlossen, „sich allein um das Mechanische, die Luftdruckkurve, zu kümmern. Die ‚physikalische Akustik‘ enthält dann also in der Tat das, was vom Schall, von Musik bleibt für einen, der taub ist“ (Wagenschein, 2009d, S. 98). Grebe-Ellis (2006a) bemerkt analog dazu, dass die in der Optik gebräuchlichen Begriffe der Wechselwirkungen (Brechung, Beugung, Streuung usw.) auf eine Reduktion auf mechanische Vorgänge zurückzuführen sind. Deren sinnliche Wahrnehmung wird im Wesentlichen durch Tasterfahrungen ermöglicht und damit die Optik als *Lehre vom Sichtbaren* zu einer *Optik ohne Auge* oder *Optik für Blinde*. Die sich daraus ergebenden Konsequenzen für die Physikdidaktik sind an anderer Stelle ausführlich formuliert (Grebe-Ellis, 2005, S. 21ff). Mit dem Bezug auf die Sinne versucht PbPU eine *Physik für Wahrnehmende* zu entwickeln. Dies setzt voraus, dass ein phänomenbasiertes Vorgehen auch den Gebrauch der Sinne fördern muss. Das genaue Beobachten mithilfe der Sinne muss immer wieder geübt und geschärft werden, um somit die qualitative Wahrnehmung zu schulen (Mackensen, 1992).

Der Aspekt der Subjektivität beinhaltet neben der Rolle der sensuellen Wahrnehmung auch ein Merkmal von PbPU, das sich als Folge daraus ergibt – die Einbindung der Schülerinnen und Schüler in das Experiment. Der Einbezug der Schülerinnen und Schüler in das Phänomen, also das Teilwerden des Experiments, soll verhindern, dass sich durch das passive, danebenstehende Betrachten eines Phänomens für sie nur eine eingeschränkte Perspektive ergibt. Schön (1985) bemerkt dazu, dass im Allgemeinen die Experimente im Physikunterricht quer zum Publikum auf dem Labortisch aufgebaut werden und somit einer zweidimensionalen Abbildung im Schulbuch entsprechen. Die

Teilnahme des experimentierenden Subjekts gewährleistet eine „dritte Dimension“ und damit eine erweiterte Beobachtungsperspektive auf das Phänomen. Schülerinnen und Schüler, die einmal selbst geschaukelt haben, erfahren mehr über wirkende Beschleunigungen und Kräfte als diejenigen, die es nur von der Seite beobachten. Die Kreiselkräfte und deren Auswirkungen erlebt man eindrucksvoller, wenn man einmal selbst ein Kreisel war. Sommer (2005) ersetzt in diesem Kontext die Begriffe *subjektiv* und *objektiv* aufgrund ihrer historisch bedingten Konnotation durch *eingebunden* und *abgelöst* und definiert eben solche Versuche, in denen der Experimentator sich als unmittelbarer Teilnehmer in den Versuch hinein begibt, als *eingebundene Versuche* (vgl. auch Grebe-Ellis, 2005, S. 39). Die Berücksichtigung von eingebundenen Versuchen im Unterricht stellt sich auch in der Expertenbefragung als ein klares Merkmal von PbPU heraus (Westphal et al., 2011). Das bedeutet jedoch nicht, dass keine *abgelösten Versuche* stattfinden, da damit der Gewinn einer „dritten Dimension“ durch den Verlust einer anderen nichtig wäre. Der ständige Wechsel zwischen beiden Perspektiven ermöglicht erst den Erkenntnisgewinn durch die eigene Beobachtung, indem z. B. in der Optik die Ansicht eines realen Bildes mit dem verglichen wird, „was ein Bildpunkt 'sieht'“ (Schön, 1985, S. 30).

Grebe-Ellis (2005) macht darauf aufmerksam, dass es offensichtlich verschiedene *Grade*, wenn nicht sogar *Arten der Einbindung* gibt. Im Bereich der Optik gibt es Versuche, in denen mehrere eingebundene Subjekte alle die gleiche Beobachtung machen, z. B. wo sich der Ort des Spiegelbildes eines vor einem ebenen Spiegel stehenden Objektes befindet. Dagegen ist der Ort des Bildes eines unter Wasser befindlichen Objektes davon abhängig, von wo das Subjekt in das Wasser hinein schaut. Hinzu kommt die Tatsache, dass die Begriffe *eingebunden* und *abgelöst* zumeist nur in der Optik verwendet werden, in der sich grundsätzlich alle Zusammenhänge als Beleuchtungsverhältnis oder Sichtbeziehung beschreiben lassen (Grebe-Ellis, 2005; Grebe-Ellis, Theilmann & Rang, 2009; Sommer, 2005). Deshalb plädieren Müller und Schön (2009) dafür, das Begriffspaar nicht voreilig auf andere Bereiche der Physik auszuweiten:

„Eingebunden“ und „abgelöst“ gehören somit nicht zum *methodischen* Vokabular einer erscheinungsorientierten Physik im Allgemeinen, sondern erscheinungsorientiert vorgehend zeigt sich ein wesentlicher Zug aller optischen Phänomene. Und dieser erlaubt stets zwei Perspektiven auf ein und dieselbe optische Erscheinung: eingebundene und abgelöste. Ob ein ähnlicher wesentlicher Zug auch in anderen Phänomenbereichen zu finden ist, kann von der Optik her gar nicht entschieden werden. [...] Die Tatsache allein, dass wir sensitive Wesen sind, erlaubt jedenfalls noch keine Schlüsse (S. 219).

Dennoch soll die subjektive Einbindung der Schülerinnen und Schüler in Versuche hier als Merkmal von PbPU festgehalten werden, da sie erstens in dem oben skizzierten Sinne auch auf andere Bereiche der Physik anwendbar ist und zweitens die Optik von allen

Bereichen des phänomenbasierten Physikunterrichts am umfangreichsten und detailliertesten ausgearbeitet ist.

Als logischer Schluss folgt aus dem Aspekt der Subjektivität, also dem Respektieren der sinnlichen Wahrnehmung und der Einbindung in die Phänomene bzw. Versuche, auch eine die Erkenntnisgewinnung betreffende Eigenschaft von PbPU. *Wissen, Erkenntnis, Sicherheit* oder *Verstehen* entsteht beim phänomenbasierten Vorgehen durch eigene Beobachtungen und beruht damit auf eigenen, autonomen Einsichten (Krüger, 2011b; Theilmann, 2008).

Aspekt der Mediation

Das von Wagenschein an viele Stellen thematisierte Problem, dass sich viele Schülerinnen und Schüler im Physikunterricht einer abstrakten und rein kognitiven Welt ausgesetzt wiederfinden, abgetrennt von ihren Alltagserfahrungen, wird von Østergaard et al. (2008) als Kern der phänomenologischen Kritik am aktuellen Physikunterricht bezeichnet. Der Aspekt der Mediation fasst die gesammelten Merkmale zusammen, die diese Kritik und somit die Vermittlung zwischen *Alltagswelt* und *Physikwelt* betreffen. Er steht damit in einem engen Zusammenhang zum Aspekt der Subjektivität, da eine der zu vermittelnden Welten, die Alltagswelt, sich durch die Erfahrungen, Wahrnehmungen und den Erlebnissen der Schülerinnen und Schüler konstruiert. Laut Expertenbefragung ist diese Vermittlung charakteristisch für PbPU.

Erb und Schön (1996) nennen als Ausgangspunkt ihrer fachdidaktischen Überlegungen für die Entwicklung ihres Optikcurriculums „die visuelle Wirklichkeit, wie sie von den Schülern erlebt, beschrieben und erklärt wird“ (S. 31). Dem liegt das von Erb (1994) formulierte Leitziel zugrunde, dass die Beobachtung die Brücke zwischen der unvoreingenommenen Wahrnehmung und der physikalischen Beschreibung darstellt und ihr deshalb eine besondere Bedeutung für die physikalische Erkenntnisbildung zukommt (Leitziel 2, S. 15). Demnach gewährleisten die Beobachtungen die geforderte Vermittlung, wodurch sie im PbPU zu einem zentralen Element werden. Dies zeigt sich auch in der Expertenbefragung, in der die Teilnehmer die ihrer Meinung nach drei wichtigsten Betätigungen im PbPU auswählen sollten. Die meistgenannte Tätigkeit ist *Beobachten* vor *Staunen* und *Erklären* (siehe Abb. 2.2).

Krüger (2011a) stellt als Leitlinie für die Gestaltung von phänomenbasiertem Unterricht heraus, dass die Beobachtungen an möglichst alltags- oder naturidentischen Anordnungen gemacht werden müssen. Das im Unterricht präsentierte Phänomen sollte nicht zu sehr präpariert sein, um auf das Wesentliche hinzuweisen, weil damit der Bezug auf die Lebenswelt der Schülerinnen und Schüler zu schnell abgebrochen wird. Muckenfuß (1995) nennt als tiefere Ursache für solche „ästhetischen Verarmungen der Lehrgeräte“ den wissenschaftlichen Reduktionismus und plädiert dafür, auf sie zu verzichten:

Die mit speziellem physikalischem Lehrgerät gewonnenen Erkenntnisse beziehen sich auf eine diffizil bearbeitete, künstlich für diesen speziellen Zweck

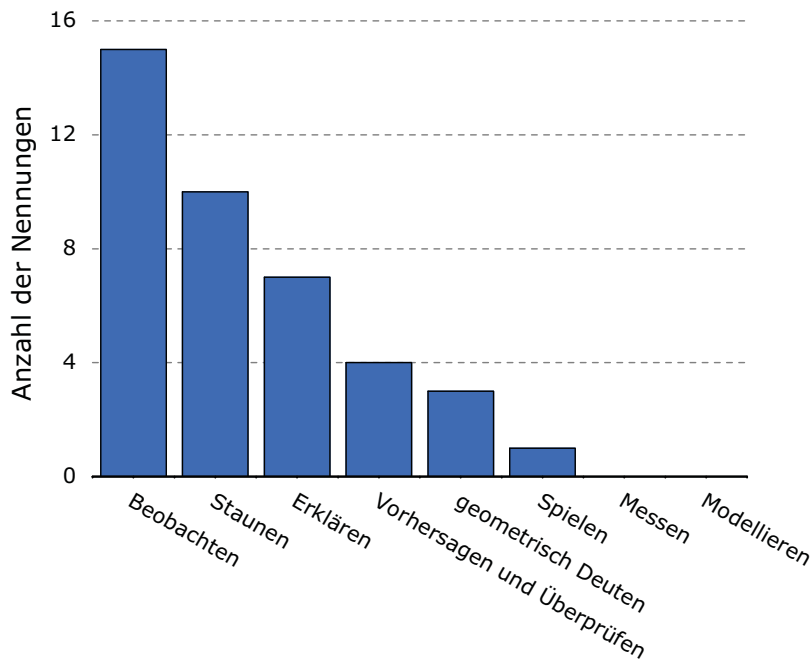


Abb. 2.2 Bedeutsame Tätigkeiten im phänomenbasierten Physikunterricht (Daten aus Westphal et al., 2011)

hergestellte „Natur“, auf künstlich produzierte Phänomene, auf die das Prädikat „physikalisch“ in einem genuinen Sinn zutrifft. Es sind von der Wissenschaft Physik *hervorgebrachte* Phänomene, keine *Naturphänomene*, aus denen die Physik zuallererst *hervorzubringen* wäre! (Muckenfuß, 1995, S. 151).

PbPU geht dieser Forderung nach und präsentiert die Erscheinungen in ihrer natürlichen Komplexität. Somit werden auch ganzheitliche Beobachtungen „neben den Phänomenen“ ermöglicht.

Das oben angesprochene Respektieren der Sinneswahrnehmung (11) kann ebenfalls als ein Merkmal von PbPU aufgefasst werden, wodurch „die Kluft zwischen Alltagserfahrungen und physikalischem Wissen zu überbrücken“ (Fedra und Schön, 1989, S. 104) versucht wird. PbPU überführt die wahrgenommene in die physikalische Wirklichkeit, auch wenn sie sich scheinbar deutlich widersprechen können (vgl. ebd.).

Die im phänomenbasierten Unterricht angestrebte enge Verzahnung von Physik mit der Lebenswelt begründet Schön (1994) mit einer These: „Damit die Physik zum aktiven Gedankengut werden kann, müssen die Schüler die physikalischen Aussagen für wahr im Sinne der Alltagserfahrungen halten können“ (S. 4). Schön merkt an, dass zwar auch Physikunterricht die Sinne anspricht, der sich nur auf die Wahrnehmung der Zeigerausschläge eines Messgerätes stützt. Jedoch entspricht die Welt des Physikraumes nur einem sehr kleinen Ausschnitt der Lebenswelt der Schülerinnen und Schüler. Pb-

PU vermittelt durch die Berücksichtigung der sinnlichen Wahrnehmung der Wirklichkeit physikalische Überzeugungen, die auch außerhalb des Physikraumes gelten. Diese Überzeugungen gleichen demnach nicht nur Glaubenssätzen, die nur (unter bestimmten Bedingungen) in der Physikwelt gelten (vgl. Buck & Mackensen, 2006; Wagenschein, 2009a). Durch den ständig möglichen, nie ganz abreißenden Rückbezug auf die Alltagssituationen nutzt PbPU die Physik als Erkenntnishilfe im Alltag (Schön, 1994).

Die Vermittlung zwischen Lebenswelt und physikalischen Konzepten ist eine Eigenschaft, die nach Østergaard et al. (2008) der phänomenbasierte Physikunterricht mit *kontext-orientiertem Physikunterricht* gemeinsam hat. Die Autoren nennen jedoch auch den entscheidenden Unterschied: PbPU betrachtet die lebensweltlichen Phänomene als Ausgangspunkte und nicht als Anschauungen oder Beispiele für das physikalische Wissen. Man sucht demnach im PbPU nicht nach geeigneten Kontexten um die Physik besser verständlich zu machen, sondern betrachtet die Phänomene der Lebenswelt als Eingang in das Naturverständnis.

Aspekt der Affektivität

Physik als Erlebnis? Diesen fragenden Titel eines Zeitschriftenbeitrags (Schön, 1985) kann man auf zwei Weisen verstehen: 1. Sollte Physik(unterricht) ein Erlebnis sein bzw. ist dies ein adäquater Weg, Physik zu lehren, weil er den Lernerfolg unterstützt? 2. Wie kann Physik(unterricht) zum Erlebnis werden? Die erste Frage soll hier nicht weiter betrachtet werden. Sie kann normativ oder auch empirisch fundiert beantwortet werden und deutet zum Teil auf die Forschungsfrage der vorliegenden Arbeit hin. Die Antwort auf die zweite Frage ist durch den phänomenbasierten Physikunterricht gegeben. Oder anders ausgedrückt: PbPU berücksichtigt die affektive Komponente des Physikunterrichts!⁷ Mackensen (1992) formuliert dies sogar noch strenger, indem er vermutet, dass PbPU der einzige Weg ist, zu einer emotionalen und affektiven Erziehung beizutragen:

Wenn es nur genügend gelänge, zu qualitativen Gedankenzusammenhängen innerhalb voller Wahrnehmungserlebnisse nach der hier vertretenden Erkenntnismethode an den Phänomenen zu kommen, *so müßte ein solcher Schulstoff, und eigentlich nur ein solcher, Gelegenheit der emotionalen, affektiven und damit letztlich sozialen Erziehung geben* [Hervorh. d. Verf.], nach der immer wieder gerufen wird (S. 7f).

Wie PbPU die Berücksichtigung dieser Komponente leistet, demonstriert Schön (1984, 1985) anhand zweier Beispiele: einem großen Hohl- und Wölbspiegel sowie einer großen Fresnellinse. Auch das Doppelschattenexperiment, ausführlich beschrieben von Weber und Schön (2000), kann als weiteres Beispiel hinzugezählt werden. Diese affektiven Versuche kommen der Forderung von Wagenschein (2009d) nach, möglichst große, instrumental einfache Demonstrationen von Phänomenen vorzunehmen, die noch vor dem

⁷ teilweise sogar explizit, siehe Untertitel von Schön (1984): „Beispiele für eine Berücksichtigung affektiver Komponenten im Physikunterricht“

Messen zum ruhigen Anschauen anregen. Der Autor nennt neben seinem bekannten großen Pendel noch „meterlange leuchtende Spektren, die Farbenspiele der sogenannten Gasentladungen, das Foucault'sche Pendel, die Gravitationswage und [...] eine große ständige Camera obscura, zum Hineingehen“ (S. 107).

Damit Physikunterricht zum Erlebnis werden kann, nennt Schön (1985) fünf Bedingungen, die auf eine Definition von *Erlebnis* und Merkmalen eines *Erlebnisobjektes* aufbauen:

1. „Aufbau und Versuchsdurchführung müssen der Bedeutung des Phänomens angepaßt sein! Zentrale Phänomene sollten durch ihre Gestaltung (Aufbau, Größe, Farbe, Material, Umfeld usw.) Gewicht bekommen.
2. Wird bei der Behandlung eines Phänomens auf eine alltägliche Wahrnehmung zurückgegriffen, sollte ihr durch Isolieren, Verfremden oder ähnliches künstlich Seltenheit oder Besonderheit verschafft werden.
3. Das Entstehen des Phänomens und seine Dynamik sollten in den Unterricht einbezogen werden.
4. Aufbau und Gestaltung des Versuches sollen den Schüler in das Phänomen einbeziehen. Dabei sind solche Versuche von besonderem Reiz, die unterschiedliche Beobachtungen auslösen.
5. Der Lehrer muß sich selbst vom „Erlebnis des Versuches“ gefangen nehmen lassen, damit der Funke des Interesses auch auf die Schüler überspringen kann“ (S. 29ff).

Die 1. Bedingung ist bereits oben in der Forderung von Wagenschein angesprochen und sie soll ein Ergriffensein bei den Schülern hervorrufen. Die 2. Bedingung erklärt sich weitestgehend selbst. In der 3. Bedingung steckt einerseits eine genetische Eigenschaft. Das Entstehen eines Phänomens oder eines Versuchs sollte für die Schülerinnen und Schüler nachvollziehbar sein, wie es z. B. bei einem schrittweisen Aufbau eines Drehspiegelversuchs der Fall wäre. Andererseits steckt darin auch eine zeitliche Komponente. Dieses Vorgehen verlangt viel Zeit, oftmals mehr, als es der Unterricht eigentlich zulässt. Jedoch gewinnt man erst so die Möglichkeit, sich in ein Phänomen zu vertiefen. Ein Unterricht, der die Sachverhalte als Ganzes betrachten möchte (einschließlich seiner unphysikalischen Aspekte), braucht Ruhe und Konzentration (vgl. Schön, 1984).

Die 4. Bedingung ist bereits oben diskutiert (Aspekt der Subjektivität) und braucht hier nicht weiter erörtert werden. Jedoch kommt unter dem Aspekt der Affektivität hinzu, dass neben der Wahrnehmung auch emotionale Einwände seitens der Schülerinnen und Schüler berücksichtigt werden (Buck & Mackensen, 2006). Die 5. Bedingung bezieht sich im Wesentlichen auf die Eigenschaften des Lehrers und weniger auf Merkmale eines speziellen Physikunterrichts und soll deshalb hier keine weitere Beachtung finden.

All diese Forderungen und Merkmale eines Physikunterrichts, der Physik zu einem emotionalen Erlebnis werden lässt, machen den Aspekt der Affektivität von phänomenbasiertem Physikunterricht aus. Die Nennungen der bedeutsamen Tätigkeiten im PbPU (siehe Abb. 2.2) unterstützt die Aussage, dass der Unterricht die Affektivität im besonderen Maße berücksichtigt. Nach *Beobachten* ist *Staunen* die meist genannte Betätigung.

Aspekt der Modellfreiheit/Modellkompetenz

Die Aussage „Phänomenbasierter Physikunterricht ist modellfrei“ wäre nicht nur sehr plakativ und vielleicht zu oberflächlich, sondern sie wäre falsch. *Phänomenbasiert* bedeutet *auf den Phänomenen stehend* und *von den Phänomenen ausgehend* (siehe Abschnitt 2.1). Der Begriff grenzt sich ab zu *modellbasiert*, was dementsprechend ein enger Bezug zu und Orientierung an physikalischen Modellen bedeuten würde. Der Begriff *modellfrei* geht auf Mackensen und Ohlendorf (1998) zurück, die in dem Themenheft *Modellfreie Optik* einen Optikkurs vorstellen, der ohne Verwendung eines Lichtmodells auskommt und somit tatsächlich modellfrei ist. Das Selbstverständnis der dahinterliegenden Konzeption der *Optik der Bilder* (Maier, 2003) ist von Grebe-Ellis (2006a) ausführlich dargestellt. Er macht zudem darauf aufmerksam, dass der Begriff *modellfrei* sich als ungeeignet erweist, weil er im Zuge einer fachdidaktischen Debatte vom eigentlichen Anliegen der *Optik der Bilder* nur ablenkt. Selbst im engen Rahmen der *modellfreien Optik* geht es nicht vordergründig darum, Modelle „krampfhaft“ zu vermeiden, sondern um „einen Erkenntnisansatz, der mit dem Anspruch verbunden ist, die Struktur reduktionistischer Erklärungen kritisch zu reflektieren und Urteilsformen zu entwickeln, welche die Erkenntnisbefähigung des einzelnen Menschen gegenüber der Natur nicht unterdrücken, sondern fördern können“ (Grebe-Ellis, 2006a, S. 2). Demnach ist es auch kein Merkmal von dem viel allgemeiner gefassten PbPU, auf „nützliche Gedankenformen“ (ebd., S. 2) wie den Modellen zu verzichten. Dies spiegelt auch die neutrale Bewertung (mit großer Streuung) der Aussage „*Im PbPU werden keine Modelle benutzt*“ in der Expertenbefragung wider (Westphal et al., 2011). Als modellfrei kann phänomenbasierter Unterricht demnach nicht bezeichnet werden. Charakteristisch für PbPU ist die Art und Weise, wie mit Modellen umgegangen wird. Dies beinhaltet, inwiefern sie kritisch reflektiert, sie zu Erklärung der Natur genutzt oder sie je nach Phänomen gewechselt werden.

Das Kernmerkmal von PbPU unter dem Aspekt der Modellfreiheit/Modellkompetenz ist die Art der Verwendung von Modellen bei der „Erklärung“ von Phänomenen. Dies könnte man sehr kurz vorwegnehmend zusammenfassen: Sie werden dafür nicht verwendet. PbPU hat am Ende jeder Erkenntnisgewinnung nicht das Ziel, zur elementaren, alles darlegenden Erläuterung in Form eines Modells durchzustoßen (Theilmann, 2006a). Genauso wenig sind im PbPU die Modelle der Ausgangspunkt, aus dem die Eigenschaften von Erscheinungen abgeleitet werden (Sommer, Grebe-Ellis & Vogt, 2004). Vielmehr wird angestrebt, auf hypothetische Annahmen über Vorgänge zu verzichten, die über das Beobachtbare hinausgehen (Maier, 2004; Theilmann, 2008), also ein „Vor-

rang des Unmittelbaren“ (siehe Untertitel von Wagenschein, 2009a, S. 96). „Vorrang“ bedeutet nach Grebe-Ellis (2005), dass „die Frage nach einem hinter den Erscheinungen liegenden, verursachenden Mechanismus *zunächst* [Hervorh. d. Verf.] zurück“ (S. 46) steht. Eine distanzierte reduktionistische Erklärung der Phänomene und die Ästhetik mathematischer Einfachheit sollen den Schülerinnen und Schülern also nicht vorenthalten werden. Jedoch könnten sie diese erst dann richtig wertschätzen, wenn sie die Gelegenheit hatten, sich mit den Phänomenen vorher richtig vertraut zu machen.

Gelingt die reine Verwendung von Konzepten, die sich auf Beobachtbares beziehen, nicht, so liegt dann bei der Verwendung von Modellen die Betonung auf der *Beschreibung* des Wahrgenommenen (Buck & Mackensen, 2006). Als *beschreibende* Modelle eignen sich besonders solche, bei denen die Gefahr einer Verwechslung mit der Wirklichkeit gering ist. Das Fermat-Prinzip für die Beschreibung von Lichtwegen (Erb, 1994, vgl.) oder der Random-Walk für die Beschreibung der Brown'schen Molekularbewegung sind Beispiele dafür, weil sie keine Entitäten voraussetzen, dessen Existenz von Schülerinnen und Schülern zum Schluss nicht mehr angezweifelt werden. Auch der Wellenformalismus, der eine Welle als rein mathematisches Konstrukt behandelt, könnte rein beschreibend interpretiert werden: „Bei bestimmten Gegebenheiten verhält sich Licht *wie* eine Welle.“ Demnach wird im PbPU nicht auf zweckmäßige oder helfende *Vorstellungen* verzichtet. Jedoch ist bei diesem Modell die Verlockung sehr groß, anzunehmen, Licht *sei* eine Welle. Dies wird dadurch begünstigt, dass in der Alltagssprache der Begriff *Welle* für real existierende Objekte verwendet wird, z. B. die *Welle am Strand* oder die *Welle im Teppichboden*. In diesem (phänomenbasierten) Sinne ist auch ein Modell ungeeignet, welches die Brown'sche Molekularbewegung mithilfe nicht beobachtbarer und somit hypothetischer Teilchen beschreibt oder sogar erklärt. Wagenschein (2009d) bezeichnet diese Gefahr der Verwechslung treffend:

Axiomatik und Deduktion bieten keinen Ausweg. Denn abstrakte Begriffe, die nicht in ihrer *Herkunft* aus den Phänomenen („genetisch“) zustande gekommen sind, werden *missverstanden*: als nicht von uns konstruierte, sondern als vorgefundene, grob materielle oder auch magische Wesenheiten, von denen man dann glaubt, dass sie als letzte Ursachen hinter allem stecken, was es gibt, und die Phänomene verursachen: das ontologische Missverständnis der Physik (S. 103).

An dieser Stelle sei jedoch angemerkt, dass bei den Aussagen „*Im PbPU werden keine Hypothesen über nicht beobachtbare Entitäten formuliert*“ und „*Im PbPU werden keine Modelle als Erklärung von Phänomenen, Effekten oder Erscheinungen benutzt*“ in der Expertenbefragung systematische Uneinigkeit zwischen den Vertretern und Lehrern auf der einen und den Externen auf der anderen Seite besteht. Die Externen stimmten beiden Aussagen nicht zu, während die anderen beiden Gruppen sie eindeutig bestätigten. Die oben getätigten Ausführungen sollten die hier gewählte Positionierung bzgl. der Merkmale von PbPU deutlich gemacht haben: Die erste Aussage ist in dieser Form zu streng formuliert. „Nicht beobachtbare Entitäten“ sind kein Tabu. Sie werden nur an den

Stellen vermieden, wo sie für eine angemessene physikalische Klärung des Sachverhalts nicht notwendig sind. Auch das hypothesenüberprüfende Experimentieren entspricht eher einem modellbasierten Vorgehen (siehe unten, Aspekt der Exploration), wird jedoch deshalb nicht komplett abgelehnt. Der zweiten Aussage wird hier zugestimmt, wenn unter „Erklärung“ das „ontologische Missverständnis der Physik“ zu verstehen ist.

Als Folge des dargestellten Umgangs mit Modellen ergibt sich als weiteres Merkmal für PbPU: die Förderung einer *Modellkompetenz* in Form von *Modellverständnis*. Leisner-Bodenthin (2006) nehmen eine Definition von *Modellkompetenz* mit der Unterteilung *deklaratives Wissen*, *prozedurales Wissen* und *Selbständigkeit* vor, wobei sich Letztere in der Umsetzbarkeit des deklarativen und prozeduralen Wissens in unterschiedlichen und unbekannten Situationen äußert. Das vom PbPU geförderte *Modellverständnis* ist dem deklarativen Wissen zugeordnet und dessen Anwendung gehört zum prozeduralen Wissen. Die *Inhalte des Modells* (*Modelleigenschaften*) und das *Anwenden des Modells* sind Teile der Modellkompetenz, die nicht durch PbPU speziell gefördert werden und daher hier nicht weiter beachtet werden. Das Modellverständnis definiert Leisner-Bodenthin (2006) wie folgt: „Die Schülerinnen und Schüler wissen, dass

- physikalische Modelle vom Menschen geschaffen werden,
 - wenn die Grenzen der direkten Wahrnehmung erreicht sind.
 - um (in ihrer Gänze) nicht beobachtbare Mechanismen/Objekte zu erklären, vorherzusagen und zu veranschaulichen.
- zur Modellentwicklung Spekulation, Intuition, Annahmen und Abstraktionen notwendig sind.
- Modelle zweckmäßig sind und nicht richtig oder falsch.
- physikalische Modelle hypothetisch und vorläufig sind.
- Modelle sich in der community durchsetzen müssen“ (S. 94).

Ein so definiertes Modellverständnis wird durch PbPU aufgrund der oben genannten Art und Weise des Umgangs mit Modellen gefördert, weil eben reduktionistische Erklärungen kritisch reflektiert werden (vgl. Zitat oben Grebe-Ellis, 2006a, 2006b). Die Förderung der Modellkompetenz wird auch durch die Experten in der Befragung bescheinigt. Laut ihnen werden im PbPU Modelle nur bewusst benutzt, reflektiert bewertet und eine distanzierte Sicht auf Modelle ermöglicht.

Ein weiteres Merkmal für phänomenbasierten Physikunterricht liegt darin, dass die verwendeten Modelle lange gültig sind und sich im Laufe der Schulzeit nicht widersprechen (Weber & Schön, 2000). Das Wissen über die Inhalte der Modelle ist dann langfristig tragfähig und es entsteht bei den Schülerinnen und Schülern nicht der Eindruck, Modelle seien je nach Sachverhalt beliebig austauschbar. Als Beispiel sei hier

das *Lichtwegkonzept* genannt, das eine von der Anfangsoptik bis in die Quantenphysik gültige Modellvorstellung liefert (Erb, 1994; Werner, 2000). Weber (2003) kann zudem zeigen, dass die Verwendung dieser tragfähigen, widerspruchsfreien und eher abstrakten Modellvorstellung vom Licht dazu beiträgt, das „ontologische Missverständnis der Physik“ zu vermeiden.

Aspekt der Exploration

Unter dem Aspekt der Exploration sind die Merkmale von PbPU zusammengefasst, die sich auf Methoden der experimentellen Erkenntnisgewinnung beziehen. Eine Abgrenzung zum modellbasierten Physikunterricht besteht darin, dass das phänomenbasierte Vorgehen einen explorativen Charakter hat, in dem „Forschen [...] als die Suche nach angemessenen Begriffen für die Erfahrungen, die wir machen[, verstanden wird]“ (Theilmann, 2008, S. 237). Im Gegensatz dazu würde die modellbasierte Erkenntnisgewinnung im Physikunterricht weitestgehend dem „Modus des Physiktreibens“ (Muckenfuß, 1995) entsprechen, in dem die physikalische Erkenntnisgewinnung theoriegeleitet erfolgt und die Beobachtungen und Experimente der Überprüfung von theoretischen Vorhersagen dienen, wie es z. B. durch Fischer, Schecker und Wiesner (2004) vertreten wird. Inwiefern eine explorative, induktive Methodik in der Fachwissenschaft Physik existent sowie gewinnbringend ist bzw. inwiefern sie als didaktische Konsequenz auch im Physikunterricht angemessen vertreten sein sollte, wurde bereits ausführlich diskutiert (Grebe-Ellis, 2005; Steinle, 2004, 2005).

Krüger (2011b) charakterisiert PbPU als explorativ, da induktive Erkenntnisprozesse ermöglicht werden. Dazu müssen mehrere Versuche eingesetzt werden, die von den Schülerinnen und Schülern als zunächst unterschiedlich eingeschätzt werden. Diese lassen jedoch gemeinsame Beobachtungsbedingungen und -ergebnisse erkennen und führen zu sogenannten *Phänomenkreisen*. Ein phänomenbasierter Physikunterricht ist somit von Beginn an viel komplexer.

Ein weiteres Merkmal besteht darin, dass zu jedem Versuch des Phänomenkreises Konditional- statt Kausalaussagen getroffen werden (siehe auch Aspekt der Modellfreiheit/Modellkompetenz). Die Beobachtungen oder Phänomene werden also mit den Bedingungen ihres Erscheinens in Form von *wenn-dann* oder *je-desto-Sätzen* in Beziehung gesetzt (Grebe-Ellis, 2005; Krüger, 2011b). Durch systematischen Vergleich der Aussagen aus verschiedenen Versuchen eines Phänomenkreises gelangt man zu Gemeinsamkeiten und damit zu sich gegenseitig stützenden Beobachtungen. Diese lassen dann den induktiven Schluss zu.

Phänomenbasierter Physikunterricht muss jedoch nicht an dieser Stelle enden. Der Grad an Genauigkeit kann qualitativ und quantitativ erhöht werden, indem durch *systematische Variationen* die wirksamen, also *notwendigen Bedingungen* und außerdem *messbare Zusammenhänge* gesucht werden, die wiederum weitere Voraussagen zulassen (Schön, 1994; Sommer et al., 2004). Letztlich können diese mathematisch *beschrieben* werden (z. B. durch Geometrisierung der Anordnung). Aus den qualitativen *wenn-dann*

und *je-desto-Sätzen* werden so quantitative Beschreibungs- und Darstellungssysteme⁸ Theilmann (2006a) nennt den gesamten Prozess der explorativen Erschließung „eine Art Integration, ein Ordnen und Einbetten oder eine systematische Durchdringung der Sachverhalte“ (S. 7).

Anhand der Expertenbefragung lassen sich außerdem folgende Aussagen bezüglich des explorativen Charakters zusammenfassen, denen überwiegend zugestimmt wurde (Grad der Zustimmung absteigend sortiert). „Im phänomenbasierten Physikunterricht

- werden Bedingungen für Erscheinungen gesucht.“
- werden unterschiedliche Phänomene geordnet und verbunden (Phänomenkreise).“
- werden auf Grundlage von Gemeinsamkeiten verschiedener Phänomene allgemeinere empirische Prinzipien/Regeln formuliert.“
- werden Erscheinungen systematisch beschrieben, ohne sie zugleich zu deuten.“
- werden empirische Regeln in Form von Konditionalsätzen aufgestellt.“
- werden systematisch ähnliche Erscheinungen oder Wenn-Dann-Beziehungen gesucht.“
- werden Ursache-Wirkungs-Zusammenhänge gesucht (z. B. durch zeitliche Reihenfolge zweier Beobachtungen).“
- wird überwiegend explorativ experimentiert.“
- werden mehr Schüler- als Demonstrationsversuche durchgeführt.“

Den Aussagen, „Im phänomenbasierten Physikunterricht

- dienen die Versuche der Überprüfung von theoretischen Vorhersagen.“
- wird überwiegend Hypothesen überprüfend experimentiert.“ und
- werden Phänomene vor allem als empirischer Nachweis theoretischer Vorhersagen eingesetzt.“

wurde überwiegend nicht zugestimmt, womit von den Experten der „Modus des Physiktreibens“ dem phänomenbasierten Vorgehen nicht unterstellt wird.

⁸Quantitative Beschreibungssysteme sind beschreibende Modelle, siehe Aspekt der Modellfreiheit/Modellkompetenz. Beispiele für Darstellungssysteme sind z. B. der Farbenkreis für die Farbenlehre (qualitativ) oder die Poincaré-Kugel für die Polarisierung (quantitativ).

Abschließende Bemerkung zum *Verstehen*

Den hier zusammengetragenen Merkmalen zum phänomenbasierten Physikunterricht liegt eine Auslegung von *Verstehen* zugrunde, die im Wesentlichen durch Wagenschein geprägt ist. Bei ihm bedeutet es zum einen im Sinne des eingangs gewählten Zitates „Stehen auf den Phänomenen“ (Wagenschein, 2009c, S. 103). Zum anderen gilt für ihn: „*Verstehen* heißt Verbinden“ (Wagenschein, 1976, S. 193). Darauf bezieht sich auch Spreckelsen (1992), für den Akte des Verstehens Akte des Ordnen sind:

Phänomene sich verstehend erschließen heißt sie einzulagern in das Beziehungsgefüge des vorgängig Verstandenen. [...] Das jeweils Neue wird mit dem Alten, schon Verstandenen, verbunden, zusammengenommen („comprendre“). Etwas verstanden haben heißt, sich etwas erklären, begründen können [...]. Die gedanklichen Prozesse dabei sind die des Aufeinanderbeziehens, Zueinander-in-Bezug-Setzens, des Vergleichens, Analogisierens (S. 63).

Diese Prozesse der Zuordnung und Einordnung bedeuten für den Autor ein „Verstehen durch Ordnen“.

Auf diese Auslegung von *Verstehen* bezieht sich auch Schön (1985) bei der Begründung für den Vorzug eindrucksvoller Experimente (Aspekt der Affektivität). Aus ihnen, und somit durch PbPU begünstigt, erhalten demnach die Schülerinnen und Schüler die Sicherheit, einen Sachverhalt verstanden zu haben, indem sie das Neue und Unverstandene mit dem bereits als richtig Anerkannten vielfältig verknüpfen.

Die Prozesse des „Aufeinanderbeziehens“, „Zueinander-in-Bezug-Setzens“, „Vergleichens“ und „Analogisierens“ sind Prozesse, die dem explorativen Vorgehen entsprechen (siehe oben). Daher ist es nicht verwunderlich, dass Sommer (2005) eine Erkenntnisbildung durch Ordnung von Erscheinungsreihen als Merkmal von PbPU nennt und das *Verstehen* im PbPU sich durch Herausarbeiten von immanenten Ordnungselementen bildet (Sommer et al., 2004). Die Förderung eines Verstehensprozesses im oben genannten Sinne kann somit als Eigenschaft von phänomenbasiertem Physikunterricht aufgefasst werden, die sich unmittelbar aus den anderen genannten Merkmalen ergibt. Dies wird auch durch die Expertenbefragung gestützt, in der sich das *Verstehen* durch das *Erkennen von Ordnungen* innerhalb von Phänomenkreisen als charakteristisch für PbPU ergibt.

Zusammenfassend lassen sich die Merkmale phänomenbasierten Physikunterrichts aller fünf Aspekte wie folgt formulieren: Phänomenbasierter Physikunterricht

- bezieht die Lernenden und deren Sinneswahrnehmung ein,
- vermittelt zwischen Lebenswelt und Physikwelt,
- macht Physik zum (emotionalen) Erlebnis,

- sucht keine Erklärungen in nicht beobachtbaren Entitäten,
- verwendet kritisch reflektierte, widerspruchsfreie Modelle,
- bevorzugt eine explorative Erkenntnisgewinnung und
- fördert ein Verstehen durch Ordnen.

2.3 Lerntheoretische Verortung

Wie oben bereits erörtert wurde, haben die zusammengetragenen Merkmale phänomenbasierten Physikunterrichts unterschiedliche Wurzeln (siehe Abschnitt 2.1). Sie führen demnach auf verschiedene erkenntnistheoretische Grundhaltungen und pädagogischen Ausgangslagen zurück. Die Entwicklung phänomenbasierter Unterrichtslehrgänge geschah somit nicht auf Grundlage einer einheitlichen Lerntheorie, aus der sich die einzelnen Merkmale als Konsequenz ergaben, um ein erfolgreiches Lernen (im Sinne der Theorie) zu fördern. Für das weitere hypothesengeleitete Vorgehen im Rahmen dieser Arbeit ist es jedoch notwendig, den Untersuchungsgegenstand, also PbPU, in den Kontext einer Lerntheorie zu verorten, um aus dieser heraus Annahmen über das Erreichen oder Nicht-Erreichen von Zielsetzungen abzuleiten.

Dies geschieht nicht mit dem Anspruch, allen Ursprüngen phänomenbasierten Unterrichts in dem Sinne gerecht zu werden, als dass die gewählte Lerntheorie auch derjenigen entspricht, in deren Kontext z. B. die phänomenologischen Lehrgänge der Waldorfpädagogik entstanden sind. Die Gründe dafür liegen jedoch nicht darin, dass diese als ungeeignet oder falsch betrachtet werden. Erstens ist es m. E. nicht notwendig, das Resultat von didaktischen Bemühungen (PbPU) aus derselben lerntheoretischen Perspektive heraus zu betrachten, zu interpretieren und zu beurteilen, aus der es entstanden ist. Zweitens gibt es keine solche Perspektive, der man alle Aspekte von PbPU zuordnen kann und drittens ist es sinnvoll, den Untersuchungsgegenstand unter dem Blickwinkel einer in den Fachdidaktiken gebräuchlichen Lerntheorie zu interpretieren, also durch die vertraute „lerntheoretische Brille“ zu schauen. Dies ermöglicht eine Vergleichbarkeit mit aktuellen Forschungsergebnissen anderer Untersuchungen, denen dieselbe Theorie zugrunde liegt.

Aus diesen Gründen wird der phänomenbasierte Physikunterricht im weiteren Verlauf aus dem Blickwinkel einer konstruktivistischen Lerntheorie betrachtet. Inwiefern die Beschreibung des Lernens im PbPU aus dieser Theorie heraus nicht nur, wie eben begründet, erlaubt, sondern auch aufgrund vieler Vereinbarkeiten geeignet ist, soll hier kurz dargelegt werden.

Mit der Wahl einer konstruktivistischen Sichtweise auf Lernprozesse ist in der vorliegenden Arbeit keine erkenntnistheoretische Position in Form eines (radikalen) Konstruktivismus verbunden. Auch wenn nach Murmann (2006) eine derartige Trennung in genau diesem Zusammenhang eine „erkenntnistheoretisch abstinente Haltung“ (S. 112)

bedeutet, so ist beides jedoch klar zu unterscheiden (vgl. Widodo & Duit, 2004). Auf einer weiteren, den Untersuchungsgegenstand betreffenden Ebene bedeutet außerdem die Wahl der Lerntheorie nicht, dass konstruktivistische Standpunkte Inhalt eines phänomenbasierten Unterrichts sind.⁹ Mit diesen Anmerkungen soll also einer Verwechslung von epistemischem und lernpsychologischem Konstruktivismus sowie einer womöglich im Physikunterricht zu vermittelnden Erkenntnistheorie vorgebeugt werden.

Als Kernannahme der konstruktivistischen Auffassung vom Lernen nennt Hasselhorn und Gold (2013) zunächst den Wissenserwerb, der als individueller Aufbauprozess im Gegensatz zu einem mechanischen Abbildungsprozess zu verstehen ist. Dabei liegt der „Fokus [...] mehr auf dem *Verstehen* als auf dem Behalten von Informationen“ (ebd., S. 66). Schon in diesem „Vorrang des Verstehens“¹⁰ liegt eine Vereinbarkeit zum PbPU, in dem das *Verstehen* als aktiver, vom Schüler vorgenommener Prozess des Ordnen angenommen wird (siehe Abschnitt 2.2, S. 23). Mit der Betonung auf die Prozesse des Wissenserwerbs wird der „lernende Mensch [...] als zielgerichtet Handelnder aufgefasst, der aktiv nach Informationen sucht, diese vor dem Hintergrund seines Vorwissens interpretiert und daraus neue Konzepte und Auffassungen über die Wirklichkeit ableitet“ (Hasselhorn und Gold, 2013, S. 66). Demnach wird das Wissen nicht passiv erworben, sondern vom Lernenden aktiv konstruiert. Das konstruierte Wissen ist also eine subjektive Repräsentation der vorliegenden Information, die sehr stark von den subjektiven Vorerfahrungen abhängt. Das bedeutet jedoch nicht, dass die Realität der Information selber angezweifelt wird. Lediglich die Konstruktion des Wissens darüber ist ein aktiver und subjektiver Prozess. Das Wissen selbst muss sich jedoch durch eine hergestellte Intersubjektivität (z. B. zu anderen Lernenden) ständig bewähren.

Diese Kernannahmen sind vollständig mit phänomenbasiertem Physikunterricht vereinbar, weil sie mit ihm in keinem Widerspruch stehen. Es lassen sich gegenteilig dazu sogar viele Übereinstimmungen zwischen dem konstruktivistischen Prozess des Lernens und den Merkmalen PbPUs erkennen. Der „aktiv, zielgerichtet Handelnde“ findet sich im PbPU als eingebunden Experimentierender, sinnlich Wahrnehmender und Physik Erlebender wieder (Aspekte der Subjektivität und Affektivität). Der Interpretation der Informationen vor dem Hintergrund seines Vorwissens (Alltagserfahrung) und der Ableitung neuer Konzepte über die „Wirklichkeit“ (Physikwelt) wird im PbPU durch den Aspekt der Mediation Rechnung getragen.

Widodo und Duit (2004, S. 237ff) identifizieren Kennzeichen, die einer konstruktivistischen Lernumgebung zugeordnet werden. In drei der fünf übergeordneten Kategorien befinden sich Kennzeichen, die sich ebenfalls mehr oder weniger direkt in den Merkmalen PbPUs wiederfinden lassen (siehe Verweise in Klammern):

⁹Dennoch gibt es für die Begründung einiger phänomenbasierter Standpunkte Bezüge zu Mach, der wiederum die Gestaltpsychologie maßgeblich beeinflusste (Erb, 1994; Erb & Schön, 1996; Schön, 1994; Wagenschein, 1976). Das gestaltpsychologische Konzept des *Lernens durch Einsicht* ist ein Vorläufer der konstruktivistischen Lerntheorie (Hasselhorn & Gold, 2013, S. 63).

¹⁰siehe dem gleichnamigen Sammelband mit Beiträgen zur Pädagogik Wagenscheins (Köhnlein, 1998)

(A) Konstruktion des Wissens ermöglichen

- Zum Denken herausfordernde Probleme anbieten (*Subjektivität, Mediation und Affektivität*)
- Mit Schülervorstellungen auf „evolutionäre“ Weise umgehen (*Mediation*)
- Mit Schülervorstellungen auf „revolutionäre“ Weise umgehen (*Mediation*)

(B) Relevanz und Bedeutung der Lernerfahrungen

- Exploration der Interessen, Einstellungen und Gefühle der Schüler (*Mediation und Affektivität*)
- Lernbedürfnisse der Schüler berücksichtigen (*Subjektivität, Mediation und Affektivität*)
- Vorgänge, Phänomene oder Beispiele aus dem Alltag berücksichtigen (*Mediation*)
- Verwendung von Materialien aus dem Alltag (*Mediation*)

(C) Soziale Interaktion

(D) Unterstützung der Schüler beim eigenständigen Lernen

(E) Wissenschaft, Wissenschaftler und wissenschaftliches Wissen

- Unterschiede in Theorien und Sichtweisen anerkennen (*Modellfreiheit/Modellkompetenz*)
- Die Rolle von Beobachtungen, Evidenzen, Hypothesen, Theorien und Gesetzen in der Wissenschaft (*Modellfreiheit/Modellkompetenz*)
- Unterschiedliche wissenschaftlichen Forschungsstrategien anerkennen (*Exploration, Modellfreiheit/Modellkompetenz*)
- Die Grenzen wissenschaftlicher Erklärungen anerkennen (*Modellfreiheit/Modellkompetenz*)

Für eine genauere Erläuterung der einzelnen Kennzeichen wird auf Widodo und Duit (2004) verwiesen. Es sei jedoch angemerkt, dass die nicht aufgeführten Kennzeichen keineswegs im Widerspruch zu PbPU stehen. Sie sind jedoch nicht für ihn charakteristisch und daher in seinen Merkmalen nicht aufgeführt. Es spricht jedoch nichts dagegen, die Unterrichtform im PbPU so zu gestalten, dass soziale Interaktion und eigenständiges Lernen (Kategorien C und D) unterstützt wird.

Es wurde dargelegt, dass die Merkmale phänomenbasierten Physikunterrichts mit einer konstruktivistischen Auffassung vom Lernen vereinbar sind und sich mit den Kennzeichen einer konstruktivistischen Lernumgebung überschneiden bzw. nicht im Widerspruch stehen. Darauf aufbauend können nun individuelle Kriterien betrachtet werden,

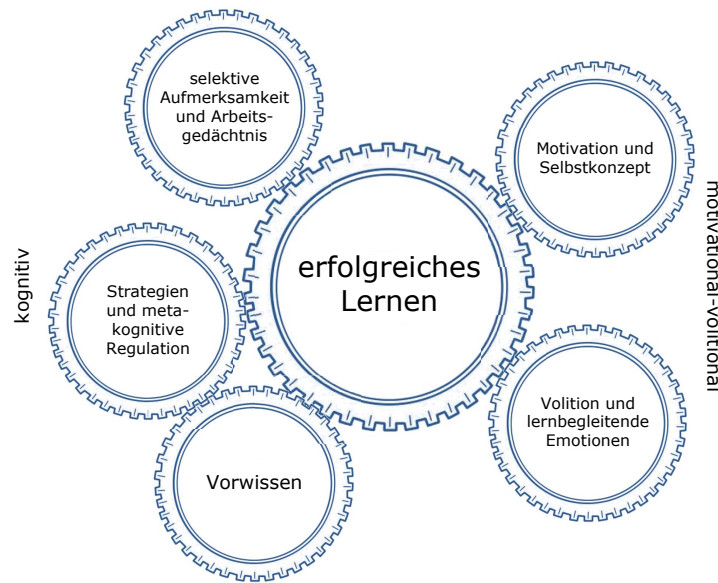


Abb. 2.3 Modell der individuellen Voraussetzungen erfolgreichen Lernens (INVO-Modell, angelehnt an Hasselhorn & Gold, 2013, S.5). Den Autoren des Modells ist bewusst, dass die Zahnräder, so wie abgebildet, mechanisch nicht lauffähig sind. Sie wollen damit verdeutlichen, dass sie bei aller Detailkenntnis noch nicht genügend über die Wirkungsweise der individuellen Voraussetzungen untereinander Bescheid wissen.

die ein erfolgreiches Lernen beeinflussen. Dies entspricht einem Wechsel von dem Fokus des Lernens im Allgemeinen auf eine differentielle Perspektive, die interindividuelle Unterschiede analysiert. Hasselhorn und Gold (2013) stellen dafür ein heuristisches Modell der **IND**ividuellen **VOR**aussetzungen erfolgreichen Lernens (INVO-Modell) zur Verfügung, das fünf Merkmalsbereiche berücksichtigt (siehe Abb. 2.3).

Für eine Verortung innerhalb dieses Modells gilt es also die Frage zu beantworten, inwiefern PbPU diesen Voraussetzungen nachkommt. Im Folgenden wird versucht, die Antwort für jeden Merkmalsbereich aus den Eigenschaften von PbPU abzuleiten, sofern dies direkt möglich ist. Dies bedeutet im Einzelnen jedoch nicht, dass diese Ableitungen empirisch fundiert sind. Tatsächlich trifft dies nur in sehr wenigen Fällen zu. Größtenteils sind die Merkmale PbPUs normative Forderungen. Wenn eine normative Forderung an den Unterricht mit den Voraussetzungen erfolgreichen Lernens vereinbar ist, so kann dem Unterricht in diesen Fällen eine Modellkonformität unterstellt werden. Inwiefern die konkreten Unterrichtsvorschläge den eigenen normativen Forderungen nachkommen und ein erfolgreiches Lernen gewährleisten, muss letztendlich empirisch überprüft werden.

Der Merkmalsbereich *Selektive Aufmerksamkeit und Arbeitsgedächtnis* geht von einem Grundmodell der Informationsverarbeitung aus, indem Lernen als Kette von Prozessen der Informationsaufnahme, -transformation und -organisation beschrieben wird. In sogenannten *sensorischen Registern* werden Informationsreize von außen für kur-

ze Zeit festgehalten, allerdings noch nicht bewusst wahrgenommen. Sobald jedoch bewusst oder unbewusst Aufmerksamkeit auf eine Auswahl aus dem Register gerichtet wird, beginnt der Lernprozess, indem diese Auswahl im *Kurzzeit-* oder *Arbeitsgedächtnis* verarbeitet wird. Die Kapazität dieses *Arbeitsgedächtnisses* ist begrenzt, weshalb es für den Lernenden wichtig ist, die für ihn relevanten Informationen im Register zu filtern (*selektive Aufmerksamkeit*). Die Kapazität unterscheidet sich von Person zu Person sehr stark, was viele Unterschiede in der Lernleistung erklärt. Allerdings hängen die Prozesse der selektiven Aufmerksamkeitszuwendung einerseits vom Vorwissen des Lernenden ab und können andererseits trainiert werden (vgl. für die gesamte Darstellung des Merkmalsbereichs Hasselhorn & Gold, 2013, S. 71ff).

Der Einfluss von Unterricht auf die Kapazität des Arbeitsspeichers ist m. E. nicht in einer relevanten Größenordnung, sofern es überhaupt durch Trainings möglich ist, sie zu erhöhen. Die in Abschnitt 2.2 (Aspekt der Subjektivität) angesprochene Schulung der sinnlichen Wahrnehmung, das Trainieren der genauen Beobachtung, das Kennenlernen von Beobachtungsstrategien, die häufige Konfrontation mit komplexen alltagsnahen Phänomenen und das Suchen nach den Bedingungen der Erscheinung sind allerdings Prozesse mit positivem Einfluss auf die selektive Aufmerksamkeit. Einerseits stellen sie das dafür wichtige Vorwissen zur Verfügung und andererseits trainieren sie, die relevanten Informationen zu filtern.

Der Merkmalsbereich *Vorwissen* ist zentral für eine konstruktivistische Lerntheorie. Je mehr (Vor-)Wissen dem Lernenden zur Verfügung steht und je besser er damit umgehen kann, desto besser sind die Voraussetzungen für ein erfolgreiches Lernen. Dazu muss es jedoch a) beim Lernprozess tatsächlich aktiviert werden und b) kompatibel mit der zur Verarbeitung stehenden Information sein. Ist der zweite Punkt nicht gegeben, sodass es zu einem Konflikt zwischen Vorwissen und der Information kommt, kann die Verstehens- und Behaltensleistung sogar behindert werden (vgl. Hasselhorn & Gold, 2013, S. 83ff). Auf beides nimmt PbPU in besonderer Form Rücksicht.

Den größten Fundus an bestehendem Vorwissen finden Schülerinnen und Schüler in ihrer Alltagswelt in Form von nicht mehr hinterfragten Grunderfahrungen. Der Aspekt der Mediation stellt zunächst sicher, dass dieses Wissen aktiviert wird, indem die Alltagswelt mit der Physikwelt verzahnt wird (siehe S. 14). Zugleich wird durch die Vermittlung zwischen beiden Welten auch die Kompatibilität des Vorwissens mit der Physik sichergestellt. Für diese Kompatibilität sorgt PbPU aber auch auf eine andere Art, nämlich durch die Verwendung widerspruchsfreier Beschreibungsformen. Jedes im Anfangsunterricht konstruierte Wissen, z. B. in Form von beschreibenden Modellen, kann als Vorwissen der Lernenden in der Sekundarstufe II aufgefasst werden. Sind diese Modelle oder Prinzipien bis in die Oberstufe widerspruchsfrei, ist die Kompatibilität des Vorwissens gegeben. Weber (2003) hat solch ein vertikal vernetztes Optikkurriculum auf seinen Einfluss auf die Qualität und Strukturiertheit des konstruierten Wissens hin untersucht. Dabei konnte er jedoch bei der Qualität keinen positiven Effekt feststellen.

Bei dem Merkmalsbereich *Volition und lernbegleitende Emotionen* soll auf den Teil der *Volition* hier nicht weiter eingegangen werden. Damit ist im Wesentlichen die Diszi-

plin und Gewissenhaftigkeit gemeint, eine motivierte Handlung durchzuführen. Sie kann damit nicht unmittelbar durch den Lerngegenstand selbst gewährleistet werden, sondern ist eher ein Persönlichkeitsmerkmal. Die *lernbegleitenden Emotionen* allerdings können sehr gut durch den Unterricht beeinflusst werden. *Emotionen* werden als mehrdimensionales Konstrukt aufgefasst, das auch affektive Komponenten (angenehmes, lustvolles bzw. unangenehmes, unlustvolles Erleben) aufweist. Es gilt als empirisch gesichert, dass Emotionen, sowohl positive als auch negative, die Erinnerungsleistung und damit das Lernen positiv verbessern (vgl. Frenzel, Götz & Pekrun, 2009). Außerdem haben sie Einfluss auf kognitive Funktionen, wie der Aufmerksamkeit (vgl. Hasselhorn & Gold, 2013). Die Berücksichtigung der Affektivität im PbPU (siehe Abschnitt 2.2) schafft Gelegenheiten für solch lernwirksame Emotionen und wirkt sich somit positiv auf den Lernerfolg aus.

Lernstrategien und Strategien der *metakognitiven Regulation* sind Techniken, die vom Lernenden bewusst angewendet werden. Sie dienen dem Zweck, die Informationen besser verarbeiten zu können und regulierend auf den Lernprozess einzuwirken. Diesbezüglich gibt es keine spezifische Eigenschaften von PbPU, die diesem Merkmalsbereich explizit nachkommen, was auch eher durch fachübergreifende Methodentrainings geschehen müsste.

An dieser Stelle sei auch angemerkt, dass hier nicht weiter auf konstruktivistische Lehrformen oder Unterrichtsgestaltungen Bezug genommen wird. Das hat zwei Gründe: Erstens lässt sich aus einer konstruktivistischen Auffassung über das Lernen nicht unmittelbar eine Lehrform ableiten, wie die Inhalte konkret zu „vermitteln“ seien. Auch bei einer darbietenden Form, wie einem Lehrervortrag, konstruiert sich das Wissen nach den oben gemachten Annahmen. Jedoch kann sich die Qualität des Wissens von demjenigen unterscheiden, welches z. B. in einer entdecken-lassenden Lernumgebung „erworben“¹¹ wurde. Bei den konstruktivistischen Ansätzen der Unterrichtsgestaltung geht man also davon aus, dass durch eine bestimmte Gestaltung der Lernumgebung das (konstruktivistische) Lernen begünstigt wird (Hasselhorn & Gold, 2013).

Zweitens ist aus den Merkmalen phänomenbasierten Unterrichts keine spezielle Lehrform ableitbar. Auch hier gibt es viele denkbare Methoden und Lehrformen, sowohl darbietende als auch entdeckenlassende usw., auf die PbPU jedoch nicht festgelegt ist. Eine weitere Betrachtung unter dem Aspekt der Lehrform wäre demnach nicht sinnvoll.

Der letzte verbleibende Merkmalsbereich der Voraussetzungen erfolgreichen Lernens ist *Motivation und Selbstkonzept*. Die Einflussnahme seitens des Unterrichts auf die Motivation der Schülerinnen und Schüler führt zu dem Konstrukt *Interesse*. Der Zusammenhang zwischen Interesse, Selbstkonzept und Lernleistung wird ausführlich in den Abschnitten 3.1 und 5.2 dargestellt. Es bleibt die Frage offen, inwiefern PbPU auf das Interesse bzw. dem Selbstkonzept einwirkt. Aus den Merkmalen lässt sich kein direkter Zusammenhang herstellen, wie es bei den normativen Eigenschaften in den ande-

¹¹ Wissen *vermitteln* oder *erwerben* entsprechen eher einer kognitivistischen Auffassung vom Lernen und Lehren.

ren Bereichen möglich ist. Jedoch kann „Interessantheit“ auch weniger ein normatives Merkmal für Unterricht sein, als eine Zielvorstellung. Dies führt direkt zu den Zielen phänomenbasierten Physikunterrichts.

2.4 Ziele phänomenbasierten Physikunterrichts

Erb (1994) stellt der Entwicklung des Lichtwegkonzepts drei Leitziele voran, bezüglich derer eine anschließende Beurteilung des Curriculums möglich sein soll:

1. „Dem Verstehen physikalischer Zusammenhänge soll ein höherer Stellenwert als dem Wissen von Fakten eingeräumt werden. Dies wird erreicht durch Erarbeiten von Beschreibungen auf einer übergeordneten Ebene.“ (ebd., S. 14)
2. „Die Beobachtung stellt die Brücke zwischen der möglichst unvoreingenommenen Wahrnehmung und der physikalischen Beschreibung dar. Ihr kommt deshalb besondere Bedeutung für die physikalische Erkenntnisbildung zu.“ (ebd., S. 15)
3. „Die wesentlichen Unterrichtsinhalte müssen über einen weiten Bereich tragfähig sein. Die Verbindungen zu anderen Unterrichtsinhalten müssen deutlich werden.“ (ebd., S. 16)

Für eine weitere Erläuterung dieser Leitziele sei auf Erb (1994) oder Weber (2003) verwiesen. Diese Ziele treffen eigentlich Aussagen über den gewünschten Endzustand des Curriculums und somit keine darüber, welche Zielvorstellung der Unterricht selbst verfolgt. Dennoch könnten aus ihnen solche Ziele von phänomenbasiertem Physikunterricht abgeleitet werden. PbPU möchte demnach ein Verstehen physikalischer Zusammenhänge herausbilden, welches über das reine Faktenwissen hinaus geht (Leitziel 1). Wie dabei *Verstehen* aufgefasst wird, entspricht den Ausführungen in Abschnitt 2.2, S. 23 und bei Erb (1994, S. 11ff). Genauso könnte Leitziel 2 so interpretiert werden, dass PbPU eine Brücke zwischen unvoreingenommener Wahrnehmung und physikalischer Beschreibung herstellen möchte. Ein nach Leitziel 3 konstruierter Unterricht führt nach Weber (2003) zu einer „Festigung des Gelernten und zu Erfolgserlebnissen [...], bei denen die Schüler die Nützlichkeit des vorangegangenen Lernens spüren“ (S. 77). Auch dies kann als ein Ziel von PbPU angesehen werden.

Ähnlich, wie bei den Leitzielen von Erb können auch die Merkmale PbPUs auf implizite Zielvorstellungen hin analysiert werden. Bei dem Aspekt der Subjektivität bedeutet dies:

1. Phänomenbasierter Physikunterricht hat das Ziel, die qualitative, sinnliche Wahrnehmung zu schulen und Sicherheit sowie Erkenntnis, beruhend auf eigenen autonomen Einsichten, zu schaffen.

Aus dem Aspekt der Mediation folgt, fast der Interpretation von Leitziel 2 gleichend:

2. Phänomenbasierter Physikunterricht hat das Ziel, zwischen der Lebenswelt der Schülerinnen und Schüler und der Physikwelt zu vermitteln.

Analog folgt aus den weiteren Aspekten z. B.: Phänomenbasierter Physikunterricht hat das Ziel,

3. ... Physik zu einem emotionalen Erlebnis zu machen.
4. ... Modellkompetenz zu fördern.
5. ... ein weitreichendes Naturverständnis ohne Erklärungen durch hypothetische Entitäten zu lehren.
6. ... ein Verstehen durch Exploration und Herausarbeiten immanenter Ordnungselemente zu fördern.

Zusätzlich zu diesen Absichten, die zugleich als Merkmale PbPUs herausgearbeitet wurden, gibt es ein häufig explizit formuliertes Ziel: *Phänomenbasierter Physikunterricht soll Interesse wecken!* Erb und Schön (1996) hoffen, dass durch den Bezug zur Alltagswelt der Abnahme des Interesses entgegengewirkt werden kann. Theilmann (2006a) sieht im erscheinungsorientierten Verständnis eine „Einladung an neues Interesse an der Welt“ (S. 8). Auch an anderer Stelle nennt der Autor als Lernziel die „Förderung des autonomen Interesses am umgebenen Naturgeschehen“ (Theilmann, 2008, S. 241). Genauso bezieht sich Grebe-Ellis (2007) auf das Interesse bei seiner Forderung nach einem Physikunterricht, der das Subjekt angemessen berücksichtigt:

Weshalb sollten sich Schülerinnen und Schüler für einen (physikalischen) Weltzugang interessieren, wenn die Sicherheit und Tragfähigkeit der (physikalischen) Erkenntnisse, die diesen Zugang bilden, darauf beruht, dass sie, diese Schülerinnen und Schüler, als erkennende und erlebende Subjekte in dieser Welt nicht vorkommen? (S. 2)

Phänomenbasierter Physikunterricht hat also das Ziel, das Interesse zur Physik und zum Physikunterricht zu fördern. Das zeigt sich einerseits in den Hintergründen, die zu den charakteristischen Merkmalen geführt haben und andererseits wird dies auch explizit als Anliegen formuliert. Somit ist es also eine Intention hinter den Gestaltungsmerkmalen PbPUs, auch dem letzten Merkmalsbereich im INVO-Modell nachzukommen und so im Sinne des Modells ein erfolgreiches Lernen zu begünstigen.

2.5 Phänomenbasierte Anfangsoptik

Nachdem bisher der phänomenbasierte Physikunterricht im Allgemeinen betrachtet wurde, soll jetzt der Blick auf die Optik eingegrenzt werden. Die Konzepte und curricularen Vorschläge zur phänomenbasierten Optik sind im Vergleich zu den anderen Themenbereichen der Physik (Mechanik, Thermodynamik usw.) am detailliertesten erarbeitet

und erprobt. Sie umfassen die Inhalte durchgängig vom Anfangsunterricht bis zur Sekundarstufe II. Der Grad an Reduktion und Unterrichtsplanung für den Einsatz in der Schule sowie der Grad an Erprobung in derselben ist für die einzelnen Inhaltsbereiche noch sehr unterschiedlich. So gibt es z. B. für die Anfangsoptik oder auch die optische Beugung Vorschläge, deren Planung bis auf die Ebene einzelner Unterrichtsstunden eingeht (Weber & Schön, 2000; Sommer, 2005). Im anderen Extremfall gibt es Themen, wie z. B. die Polarisation, die zwar phänomenologisch erschlossen sind, jedoch für die Schule noch nicht weiter ausgearbeitet wurden (Grebe-Ellis, 2005). Dazwischen gibt es ein breites Spektrum an unterschiedlich differenzierten Planungen für verschiedene Unterrichts- und Schultypen (z. B. Epochenunterricht an Waldorfschulen).

Für den weiteren Verlauf der vorliegenden Arbeit wird der Fokus hier weiter eingengt auf die *phänomenbasierte Anfangsoptik*. Die Wahl auf diesen Inhaltsbereich lässt sich folgendermaßen begründen: Erstens ist die Unterrichtssequenz planerisch detailliert ausgearbeitet. Zweitens ist sie vom gesamten Umfang an Berliner Regelschulen durchführbar. Die Dauer entspricht dabei etwa der eines Pflichtmoduls der Optik im Berliner Rahmenlehrplan (Senatsverwaltung für Bildung, Jugend und Sport Berlin, 2006, P4 7/8). Auch die Inhalte der phänomenbasierten Anfangsoptik (z. B. Lichtwege, Schatten ausgedehnter Leuchten, Spiegelwelt) sind lehrplankonform, da sie bei der Erarbeitung des Berliner Rahmenlehrplans speziell berücksichtigt wurden. Drittens ist die Durchführung der Unterrichtssequenz erprobt. Sie wird von Lehrerinnen und Lehrern im näheren Umfeld der AG Didaktik der Physik an der Humboldt-Universität zu Berlin seit mehr als zehn Jahren regelmäßig unterrichtet. Der Erfahrungsaustausch ermöglichte ständige Verbesserungen und deren Umsetzung im Schulalltag. Teilbereiche der Anfangsoptik werden außerdem seit zehn Jahren im UniLab Schülerlabor als Module angeboten. Viertens gibt es einen großen Fundus an Unterrichtsmaterialien. Das sind zum einen Arbeitsblätter, Experimentiermaterialien sowie Aufgaben und zum anderen gibt es ein Schulbuch, welches sinnvoll als Begleitmaterial eingesetzt werden kann (Mikelskis, Schön & Wilke, 2006). In dem Kapitel zur Optik sind die grundlegenden Ideen, Experimente, Merksätze und Prinzipien berücksichtigt, die bei der Entwicklung der phänomenbasierten Unterrichtssequenz ausschlaggebend waren. Fünftens eignet sich die phänomenbasierte Anfangsoptik sehr gut, um sie exemplarisch für PbPU zu untersuchen. Das liegt vor allem daran, dass sie die oben genannten Merkmale phänomenbasierten Physikunterrichts weitgehend aufweist, was weiter unten nach genauerer Betrachtung der Inhalte detaillierter dargestellt ist.

Mit *Anfangsoptik* ist hier der erste Kontakt der Schülerinnen und Schüler mit optischen Phänomenen und Gesetzmäßigkeiten im Rahmen des *Physikunterrichts* in Berlin in der Sekundarstufe I gemeint. Die Inhalte könnten in anderen Bundesländern auch im Rahmen des naturwissenschaftlichen Unterrichts in der Grundstufe (Klassen 5/6) unterrichtet werden. Die Themen können grob in die Bereiche *Licht und Schatten*, *Spiegelwelt* und (*optische*) *Hebung* unterteilt werden (siehe Abb. 2.4). Die beiden Letzten werden



Abb. 2.4 Themenübersicht der phänomenbasierten Anfangsoptik. Die Reihenfolge der Themen ist nicht zwingend von links nach rechts zu verstehen.

üblicherweise mit *Reflexion (am ebenen Spiegel)* und (*optische*) *Brechung* bezeichnet.¹² Die Reihenfolge der übergeordneten Themenbereiche muss nicht zwingend eingehalten werden, wie in Abb. 2.4 vorgegeben. Krüger (2011a) schlägt z. B. einen Unterrichtsgang vor, in dem die Spiegelwelt als Einstieg in die Optik verwendet wird. Allerdings befinden sich die Themen-Kacheln (Abb. 2.4 unten) innerhalb eines Bereichs in einer sachlogischen Reihenfolge. Dies führt m. E. am sinnvollsten zu dem hier dargestellten Ablauf.

Die Bezeichnung der Kachel, welche die Einführung in die Optik darstellt, ist bewusst an die gleichlautende Episode von Wagenschein (2009b) angelehnt. Es werden demnach zunächst die Zusammenhänge von Beleuchtungssituationen erarbeitet. Dies kann sehr ausführlich durch eine „Dunkelstunde“ geschehen, wie sie von Heinzerling (1995) geschildert wird. Als wichtiges Resultat gewinnen die Schülerinnen und Schüler die Einsicht, dass Licht für das Sehen eine notwendige Bedingung und Schatten/Schattierungen für die Wahrnehmung von Formen und Gestalten ebenso wichtig ist. Die Erkenntnis, dass das Licht selbst nicht sichtbar ist, sondern nur beleuchtete oder selbst leuchtende Gegenstände und Flächen, kann als Argument genutzt werden, sich zunächst um das Wesen des Lichts nicht weiter zu kümmern. Stattdessen wendet man sich eher dem zu, was man sieht, die Ansichten (Bilder).

Der weiße Schatten dient als motivierendes Einstiegsexperiment, welches im Konflikt zu den allgemein bekannten schwarzen Schatten steht. Das liegt jedoch nicht daran, dass weiße Schatten in der Lebenswelt der Schülerinnen und Schüler nicht vorkommen. Ist man sich einmal den Erscheinungsbedingungen bewusst geworden, kann man sie häufig aufgrund der zahlreich vorkommenden ausgedehnten Lichtquellen (z. B. Leuchtstoffröhren) wiederfinden. Für die Klärung des weißen Schattens lernen die Schülerinnen und Schüler eine Experimentiermethode kennen, die sie in allen Bereichen der Optik begleiten wird – das *Prinzip Ameise* (vgl. Heinzerling, 1995; Weber & Schön, 2000). Es entspricht der eingebundenen Perspektive (siehe Aspekt der Subjektivität), welche man

¹²Beide Begriffe beziehen sich auf einen Prozess, den mechanistische Modellvorstellungen vom Licht vollziehen (vgl. Grebe-Ellis, 2006a). In einem *streng* phänomenbasierten Physikunterricht könnte auf sie verzichtet werden. Da dies jedoch nicht zwingend gefordert ist, werden sie zur besseren Verständlichkeit hier mit aufgeführt.

sich als Sichtperspektive einer Ameise vorstellen kann, die sich am Ort des (reellen) Bildes befindet. Im vorliegenden Fall ist das reelle Bild eine Schattenprojektion an der Wand. Durch den Vergleich der Ansichten der Ameise mit den Schattenbildern, also durch ständigen Wechsel zwischen eingebundener und abgelöster Perspektive, kommt man schließlich zu dem alles entscheidenden Zusammenhang zwischen der Beleuchtungssituation und der Sichtbeziehung:

Eine Fläche erscheint umso heller, je mehr Leuchten oder helle Flächen von dort aus gesehen werden können. Eine Fläche erscheint uns umso dunkler, je weniger Leuchten oder helle Flächen von dort aus gesehen werden können. [...] Es ist nur dort ganz dunkel, von wo aus weder eine Leuchte noch irgendein beleuchteter Gegenstand zu sehen ist (Mikelskis et al., 2006, S. 87).

Im Anschluss daran werden systematisch die Schattenprojektionen verschieden geformter Lampen und Schattengeber untersucht und in eine übersichtliche Darstellung gebracht. Die Form der Projektion erweist sich als abhängig von dem perspektivischen Größenverhältnis von Lampe zu Schattengeber aus der Sicht der „Ameise“ und sie enthält Informationen über die Form von beiden. Im Extremfall eines im Verhältnis zur Lampe sehr kleinen Schattengebers oder eines großen Schattengebers mit kleinem Loch erinnert die Form des Schattens an die der Lampe. Damit ist der Schritt zur Lochkamera nur noch sehr klein. Die konsequente Interpretation der projizierten Schatten als Bilder und die Erklärung ihrer Helligkeitsverläufe durch das Prinzip Ameise, lassen es zu, dass auch das Lochkamerabild als „Schattenbild“ interpretiert werden kann, dessen Gestalt (kopfstehend, seitenverkehrt), Helligkeit und Farbe aus der eingebundenen Perspektive unmittelbar nachvollziehbar und damit plausibel wird. Dafür eignet sich am besten eine große begehbare Lochkamera, z. B. ein komplett abgedunkelter Klassenraum mit kleinem Loch in der Verdunkelung am Fenster.

Das Doppelschattenexperiment bildet m. E. das Zentrum des Curriculums zur phänomenbasierten Anfangsoptik. Es wird ausführlich von Weber und Schön (2000) vorgestellt. Zentral ist es einerseits, weil es im zeitlichen Ablauf in etwa die Mitte markiert. Darüber hinaus bildet es auch gleichzeitig eine Brücke vom Thema *Schatten* zum *Spiegel*, zu dessen Gesetzmäßigkeiten die am Experiment entstandenen Fragen führen. Zu dessen Beantwortung wird jedoch das methodische Wissen zum vorherigen Thema benötigt. Außerdem weist das Doppelschattenexperiment sehr viele charakteristische Merkmale PbPUs auf.

Die Spiegelwelt wird als optisch real wirkende „Welt“ kennengelernt, in die der Spiegel wie ein Fenster die Einsicht ermöglicht. Die Dinge darin sind somit nicht ertastbar, sondern nur sichtbar. Die reale optische Wirkung wird durch Schatten verdeutlicht, die sich, durch eine in der Spiegelwelt befindliche Lampe in die reale Welt fortsetzen und umgekehrt. Das Prinzip Ameise lässt sich für die Klärung von Helligkeitsunterschieden am Spiegel genauso gut einsetzen, wenn die im Spiegel gesehenen Leuchten mit berücksichtigt werden. Mithilfe von Schattengrenzen, die als ausgewählte Sicht- und

Lichtwege interpretiert werden, wird in einer Reihe von Schülerversuchen das 1. Spiegelgesetz erarbeitet: „Gegenstand und Spiegelbild sind gleich weit von der Spiegelebene entfernt. Die Verbindungslinie zwischen Gegenstand und Spiegelbild steht senkrecht auf der Spiegelebene“ (Mikelskis et al., 2006, S. 95). Aus weiteren Experimenten lässt sich sehr leicht aus dem 1. das 2. Spiegelgesetz ableiten, welches mit dem üblichen Reflexionsgesetz identisch ist. Die Gleichheit von Reflexions- und Einfallswinkel ist in diesem Gang demnach kein durch Messungen induktiv gefundenes Gesetz, sondern unmittelbare Folge aus den geometrischen Beziehungen zwischen Gegenstand und seinem Spiegelbild. Eine ausführliche Darstellung der Überlegungen zur Spiegelwelt und auch zum 3. Spiegelgesetz findet man bei Schön (1994).

Der letzte Themenbereich der phänomenbasierten Anfangsoptik geht von den Ansichten beim Blick ins Wasser aus. Nach systematischer Ordnung der Erscheinungen aus verschiedenen Experimenten wird die Erkenntnis gewonnen, dass beim Blick ins Wasser die Dinge im Wasser zur Oberfläche hin lotrecht angehoben erscheinen. Das Maß der Hebung (Brechzahl) wird sowohl im senkrechten als auch beim schrägen Blick ins Wasser quantifiziert, indem es geometrisch vermessen wird. Theoretisch wäre damit eine physikalische Betrachtung der Erscheinungen getätigt, die äquivalent zu der üblichen Behandlung der Brechung zu verstehen ist, da beide denselben physikalischen Sachverhalt nur aus unterschiedlichen Perspektiven angehen (vgl. Grebe-Ellis et al., 2009). Bei der Hebung gewinnt man Aussagen über die (virtuellen) Bilder der unter Wasser befindlichen Objekte und bei der Brechung werden Aussagen über das Verhalten eines Modells vom Licht an der Grenzfläche getätigt. Eine phänomenbasierte Betrachtung der Brechung wären Aussagen über Lichtwege, die durch Schattengrenzen am Übergang Luft/Wasser sichtbar gemacht werden können. Auch das Prinzip Ameise wäre denkbar, ohne dafür sich ins Wasser begeben zu müssen, z. B. mit einem kugelförmigen Spiegel. Jedoch eignet sich gerade dieses Thema, um die Umkehrbarkeit der Lichtwege anzusprechen und sich nur noch gedanklich in die Ameise hinein zu versetzen. An dieser Stelle sei jedoch angemerkt, dass genau dieser Wechsel von eingebundener zu abgelöster Perspektive und umgekehrt keineswegs trivial ist. Das In-Beziehung-Setzen der Sichtwege bei der Hebung mit den Lichtwegen bei der Brechung führt sehr schnell zu Verwirrungen und Verwechslungen und ist selbst für Physik-Studierende eine sehr anspruchsvolle Aufgabe.

Die Wurzeln für den hier grob dargestellten Unterrichtsgang sind zu großen Teilen die didaktisch motivierten Ausarbeitungen von Mackensen (1992) bzw. Mackensen und Ohlendorf (1998). Diese wiederum stehen in einem engen wechselseitigen Zusammenhang mit den wissenschaftlichen Arbeiten von Maier (2003, 2004), die einen weniger didaktischen Hintergrund haben und eher grundlegender Natur sind. Maier verfolgt mit seiner *Optik der Bilder*¹³ das Anliegen, auf Mechanismen zu verzichten, die im „Verborgenen“ die Erscheinungen verursachen, was im Wesentlichen einer *modellfreien Op-*

¹³siehe Titel des Buches (Maier, 2003)

tik entspricht. Als Konsequenz betrachtet er alle optischen Erscheinungen als *Bilder*. Damit geht die Aufwertung der eingebundenen Perspektive einher (siehe S. 11), da durch sie die üblicherweise von der Seite betrachteten Beleuchtungsbeziehungen als *Sichtbeziehungen* realisiert werden (vgl. Grebe-Ellis, 2006a; Müller & Schön, 2009). Die Arbeiten von Mackensen und Ohlendorf stellen die Anwendung der *Optik der Bilder* Maiers auf den Physikunterricht an Waldorfschulen dar.¹⁴ Auf Maier, Mackensen aber auch Julius (1984) lassen sich viele der für die phänomenbasierte Anfangsoptik grundlegenden Konzepte und Überlegungen zurückführen. Dazu gehören z. B. die Einbindung in optische Versuche (Prinzip Ameise) als themenübergreifende Experimentiermethode, der photometrische Grundsatz „Hell ist es, von wo aus Helles zu sehen ist“ (Grebe-Ellis, 2006b, S. 2), die Interpretation von Schatten als Bilder und die Spiegelwelt.

Die Anregungen aufgreifend bettet Schön (1984, 1985, 1994) die Weiterentwicklungen der Anfangsoptik in das Curriculumprogramm *Vom Sehen zur Optik* ein, an die sich die Mittel- und Oberstufenoptik als *Lichtwegkonzept* mithilfe des Fermat-Prinzips und des Zeigerformalismus anschließen (Erb, 1994; Werner, 2000). Im Rahmen des BLK¹⁵-Programms SINUS¹⁶ entsteht in enger Zusammenarbeit mit den Projektschulen und den teilnehmenden Lehrern schließlich ein bis in die Feinplanung ausgearbeitetes Curriculum inkl. Lehrmaterialien (vgl. Weber & Schön, 2000). Diese Unterrichtssequenz (im Wesentlichen entspricht sie der Darstellung in Abb. 2.4) dient als Orientierung für die hier betrachtete *phänomenbasierte Anfangsoptik* und damit für den Untersuchungsgegenstand der vorliegenden Arbeit.

Nachdem die Kernelemente des Optik-Curriculums skizziert wurden, ist noch offen, inwiefern dieser Unterricht auch als phänomenbasiert bezeichnet werden kann. Wo finden sich die Merkmale PbPUs hier wieder? Oder von einem anderen Standpunkt aus gefragt: Wie sehen die bisher allgemein formulierten Merkmale PbPUs an konkreten Inhalten im Unterricht umgesetzt aus?

Der Aspekt der Subjektivität ist im Falle der Optik offensichtlich. Mit der Orientierung an einer *Optik der Bilder* ist die Rolle der wahrnehmenden Schülerinnen und Schüler aufgewertet, da es schließlich um deren Ansichten geht. Diesen Ansichten wird auch der Grad an Realität zugestanden, die beobachtbar sind. Ein weißer Schatten (siehe Abb. 2.5) wird erst einmal als ein solcher akzeptiert, ohne nach dem Grund für eine vermeintliche „optische Täuschung“ zu suchen. Die Spiegelwelt wird als optisch real wirkende Welt kennengelernt, weil sie und die von ihr verursachten Erscheinungen eben sichtbar sind. Damit geht einher, dass sich ein Bewusstsein für solche *Schwelten* bildet, deren Gegenstände nicht anfassbar sein müssen, aber dennoch optisch reale Wirkungen hervorrufen. Für das Verständnis von virtuellen Bildern, z. b. bei Linsen, kann dies sehr hilfreich sein. Ermöglicht wird die Schwerpunktverschiebung hin zu den Sichtbeziehungen

¹⁴so auch ein Untertitel: *Phänomenologischer Physikunterricht, entwickelt aus der Praxis und Theorie der Waldorfschule* (Mackensen, 1992)

¹⁵Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung

¹⁶Modul 5 – Zuwachs von Kompetenz erfahrbar machen: Kumulatives Lernen



Abb. 2.5 Der weiße Schatten

gen durch die Einbindung der Schülerinnen und Schüler, die in der phänomenbasierten Anfangsoptik z. B. durch das *Prinzip Ameise* geschieht. Aber auch der Blick in den Spiegel und in das Wasser sind Formen der Einbindung in Versuche, die üblicherweise von der Seite betrachtet werden.

Der Grad an Subjektivität in der phänomenbasierten Anfangsoptik hängt stark mit deren vermittelnder Wirkung zwischen Alltags- und Physikwelt zusammen. Die Entwicklung der physikalischen Erkenntnis erfolgt aus der Sicht der Schülerinnen und Schüler heraus und damit bezogen auf sie und ihre Wahrnehmungswelt. Das Für-Wahrnehmen der eigenen Sinnesempfindung (Spiegelwelt) trägt ebenso zu der Vermittlung bei, wie die Betrachtung alltagsnaher Gegebenheiten. Letzteres trifft z. B. bei der Behandlung von Schattenbildern ausgedehnter Lichtquellen zu, die in der Lebenswelt der Schülerinnen und Schüler häufiger vorkommen, als (ideale) Punktlichtquellen. Die phänomenbasierte Anfangsoptik konstruiert damit ein physikalisches Gebäude, in dem sowohl die Schülerinnen und Schüler selbst als auch die Gegenstände ihrer Lebenswelt vorkommen dürfen.

Dem Aspekt der Affektivität wird an mehreren Stellen Rechnung getragen, was auch gerechtfertigt ist, aufgrund der Tatsache, dass es sich um Anfangsunterricht in den unteren Klassenstufen der Sekundarstufe I handelt. Als Beispiele für Elemente, die besonders zum Staunen, Verharren und zum Äußern von Emotionen einladen, können die oben genannte Dunkelstunde, der weiße Schatten und auch das Doppelschattenexperiment genannt werden (vgl. Heinzerling, 1995; Weber & Schön, 2000). Das Spielen mit der Dunkelheit und die sorgfältige Inszenierung der Details gibt den relevanten Erscheinungen die Geltung, die ihnen zusteht, was von den Schülerinnen und Schülern auch gewürdigt wird. Die von Schön (1985) formulierten Bedingungen dafür, dass der

Physikunterricht zum Erlebnis wird (siehe Seite 17), werden in der Anfangsoptik alle berücksichtigt.

Für die erste und dritte Bedingung sei auf die Schilderung von Weber und Schön (2000) verwiesen, in der sich beide unmittelbar wiederfinden lassen. Die zweite Bedingung kann exemplarisch am weißen Schatten erläutert werden (siehe Abb. 2.5). Da es sich bei Schatten um durchaus alltägliche Erscheinungen handelt, wird als Schattengeber zunächst kein geometrisch einfach beschreibbares Objekt gewählt, sondern beispielsweise ein Würfelkantengerüst. Die Schattenprojektion ist dadurch schon bei der Verwendung einer Punktlichtquelle verfremdet und selten. Setzt man den Schattengeber zusätzlich in leichte Rotation, gibt es bei der Schattenprojektion viele eigenartige perspektivische Verzerrungen und Kippbilder. Bei der Verwendung einer Stableuchte verschwindet die Verwirrung zunächst, da die Schattenprojektion unscharf und kontrastlos erscheint. Hält man nun eine weitere Blende vor die Stableuchte, z. B. zwei Finger, erscheint inmitten dieses unscharfen Schattenbereichs eine helle Schattenprojektion des Würfelkantengerüsts, die an Schärfe vergleichbar ist, mit der dunklen Projektion der Punktlichtquelle. Durch Bewegungen der abdeckenden Finger entlang der Stablampe bewegt sich auch der weiße Schatten, was es erleichtert ihn zu identifizieren.¹⁷ Beim weißen Schatten lassen sich auch die gleichen Verzerrungen und Kippbilder wie bei seiner dunklen Entsprechung finden. In Abb. 2.5 sind einige der weißen Kanten nicht parallel und das Bild kippt zwischen zwei Würfeln, auf die man entweder von schräg unten oder schräg oben schaut.

Durch diese umfangreichere Schilderung eines Versuchs in der phänomenbasierten Anfangsoptik soll deutlich werden, wie durch Verfremdung und Einbeziehung der Dynamik des Phänomens der Physikunterricht erlebnisreicher wird. Auch die Verwendung von Kerzen als Punktlichtquellen in Schüler- und Demonstrationsversuchen und die Einbeziehung der Schülerinnen und Schüler in die Phänomene (siehe 4. Bedingung auf Seite 17) tragen zur Affektivität bei, sodass m. E. dieser Aspekt PbPUs im besonderen Maße bei der Anfangsoptik auftritt.

Der Aspekt der Modellfreiheit/Modellkompetenz ist insofern erkennbar, als dass Licht als Entität nicht mehr Bedeutung beigemessen wird, als für die Klärung der Erscheinung notwendig ist. Eine Modellvorstellung, die über den Lichtweg hinaus geht, wird nicht benötigt, womit über das Wesen des Lichts jedoch keine Aussage gemacht wird. Die Einführung der Lichtwege ist auch nur an den Stellen notwendig, bei denen auf die üblichen Darstellungen des Reflexionsgesetzes (2. Spiegelgesetz) und der Brechung hingearbeitet werden soll. Für beide physikalischen Sachverhalte gibt es eine äquivalente Formulierung, die von den Sichtwegen ausgeht (1. Spiegelgesetz und Hebung), womit keine Erklärungen in den Modellen gesucht werden müssen. Für die „Erklärung“ von Schattenformen wird ebenfalls kein Lichtstrahlenmodell benötigt, mit

¹⁷Dieser Umstand macht es auch schwer, Fotos vom weißen Schatten zu erzeugen. Die Erscheinung ist dynamisch weitaus effektvoller als in Abb. 2.5. Auch plötzliches Positionieren und Wegziehen der Finger vor der Lampe erleichtert das Auffinden der Projektion.

dem der Umgang mit ausgedehnten Lichtquellen auch umständlich wäre. An seine Stelle tritt das Prinzip *Hell ist es, von wo aus Helles zu sehen ist*, welches sich für zukünftige Unterrichtsinhalte als tragfähig erweist.

Als Beispiel für explorative Phasen, in denen durch Systematik und Ordnung ein Verständnis für einen physikalischen Sachverhalt entsteht, kann die Erarbeitung der Schattenformen ausgedehnter Lichtquellen genannt werden. Aus der Variation der Formen von Lampe und Schattengeber und einer übersichtlichen Darstellung werden die die Schattenform bedingenden Faktoren gefunden und in Form von Konditionalaussagen festgehalten. Die Erfahrungen aus zahlreichen komplexen Variationen führen letztendlich zu der Gewissheit, dass Schattenbilder Informationen aus Lampe und Schattengeber enthalten, wodurch am Ende die Schülerinnen und Schüler auch in der Lage sind, die Lampenform anhand des Schattenbildes und der Kenntnis über die Form des Schattengebers vorherzusagen. In den Konstellationen von Lampe, Schattengeber und Schattenprojektion, wo die Gewissheit der Schülerinnen und Schüler auf die Probe gestellt wird, weil der Zusammenhang zu magisch erscheint (z. b. wenn der Schattengeber sehr klein ist und plötzlich die Lampenform in der Schattenprojektion zu sehen ist), ist mit dem Prinzip Ameise immer ein Erklärungsansatz möglich. Schließlich führt dies auch zu einem Verständnis der Lochkamera ohne Lichtstrahlenmodell. Die Interpretation des Bildes als Schatten und des Gegenstandes als komplexe ausgedehnte Lampe (wenn auch nicht selbstleuchtend) fügt die Lochkamera-Abbildung nahtlos in die zuvor erarbeitete Systematik und Ordnung bzw. deren übersichtliche Darstellung ein. Rätselhaft bleibt dann nichts mehr, schließlich sehen die Schülerinnen und Schüler vom Ort der Projektionsfläche durch das Loch der begehbaren Kamera genau den Gegenstand in Form und Farbe, wie er dort auch abgebildet ist – zusammengefasst ein *Verstehen durch Ordnen*.

2.6 Bedeutung für die vorliegende Arbeit

Phänomenbasierter Physikunterricht bedingt eine besondere Form der Naturerschließung und Unterrichtsgestaltung, die sich in zahlreichen charakteristischen Merkmalen niederschlägt. Die dahinter liegenden Grundgedanken und Forderungen an Physikunterricht greifen auf eine lange physikdidaktische Forschungstradition zurück, die sich jedoch im heutigen Physikunterricht nur wenig widerspiegelt. Im Laufe der Zeit wurden jedoch viele konkrete Unterrichtsvorschläge in allen Themenbereichen der Physik entwickelt, die diesen Forderungen versuchen nachzukommen. Auch wenn der hier als *phänomenbasiert* bezeichnete Unterricht nicht aus einer abgeschlossenen didaktischen Theorie entstanden ist, sondern als Konsens verschiedener Konzeptionen mit unterschiedlichen Wurzeln definiert ist, so lässt er sich dennoch aus dem Blickwinkel einer in der fachdidaktischen Forschung üblichen Lerntheorie betrachten. Die zumeist normativ formulierten Merkmale PbPUs kommen zum großen Teil den Voraussetzungen erfolgreichen Lernens nach. Eine Voraussetzung, nämlich *Interesse* am Lerngegenstand

und -geschehen, kann nicht direkt aus den Merkmalen abgeleitet werden. Zugleich ist die Förderung von Interesse ein häufig implizit und explizit formuliertes Ziel von PbPU. Eine empirische Überprüfung dessen, ob eine phänomenbasierte Vorgehensweise das Interesse positiv beeinflusst und damit ein erfolgreiches Lernen fördert, liegt derzeit noch nicht vor.

Auf diese offene Frage hin soll stellvertretend für PbPU im Rahmen der vorliegenden Arbeit ein phänomenbasiertes Curriculum für die Anfangsoptik untersucht werden. Der dadurch entstehende Verlust an Allgemeingültigkeit wird kompensiert durch die langjährigen Erfahrungen und der damit einhergehenden Güte in der Ausarbeitung des Unterrichtskonzepts. Außerdem eignet sich dieses Beispiel für eine stellvertretende Untersuchung, da die Anfangsoptik besonders viele Merkmale PbPUs aufweist und somit als repräsentativ dafür angesehen werden kann.

Die phänomenbasierte Anfangsoptik stellt somit den Untersuchungsgegenstand der vorliegenden Arbeit dar, die auf das Ziel der Interessenförderung als Voraussetzung für erfolgreiches Lernen empirisch überprüft werden soll.

Assoziationen zum Schulfach Physik

Zwei in Abschnitt 2.4 formulierte Ziele von phänomenbasiertem Physikunterricht bestehen darin, das Interesse zu wecken und einen bruchlosen Übergang von der Alltagswelt der Schülerinnen und Schüler zum Physikunterricht herzustellen. Durch solch einen bruchlosen Übergang soll die Möglichkeit einer Identifizierung mit dem Fach gefördert werden, indem deutlich wird, dass die erlernten Gegenstände des Physikunterrichts (Physikwelt) unmittelbar in Beziehung stehen mit den alltagsrelevanten Dingen der Welt, in der die Schülerinnen und Schüler leben (Alltagswelt). Diese Identifizierung mit dem Fach entspricht der Entstehung eines positiven Zusammenhanges zwischen dem Selbst der Schülerinnen und Schüler und Physik, also einer Integration in das Selbstkonzept. In Hinblick darauf werden in dem folgenden Abschnitt zunächst das der Arbeit zugrunde liegende Interessenkonstrukt und diesbezügliche empirische Befunde vorgestellt. Des Weiteren wird ein Zusammenhang zum Image des Schulfaches und den damit einhergehenden Assoziationen zum Physikunterricht hergestellt. Schließlich soll deutlich werden, warum es sinnvoll ist, bei den beiden genannten Zielsetzungen die Assoziationen in Betracht zu ziehen.

3.1 Interesse an Physik

Das in der vorliegenden Arbeit verwendete Interessenkonstrukt entspricht der von Krapp (1992a) formulierten *Person-Gegenstands-Konzeption des Interesses*. Demnach lässt es sich auf folgende Weise definieren:

Definition (Interesse): Interesse bezeichnet die *Beziehung einer Person zu einem Gegenstand*, die sich durch eine *gefühlsbezogene Komponente* (der Gegenstand wird von der Person mit positiven Gefühlen assoziiert), einer *wertbezogenen Komponente*

(der Gegenstand hat für die Person eine besondere Bedeutung) und einer *intrinsischen Komponente* (die Person empfindet ihr Interesse nicht von außen aufgezwungen) kennzeichnen lässt (vgl. Krapp, 1999).

Der Begriff *Person-Gegenstands-Konzeption* hebt bereits eine wichtige Eigenschaft von Interesse hervor, nämlich die Gegenstandsspezifität. Eine Person hat demnach Interesse an oder für etwas, jedoch hat sie nicht Interesse im Allgemeinen. Als *Gegenstände* gelten „Sachverhalte in der Lebenswelt eines Menschen, über die Wissen erworben und ausgetauscht werden kann“ (Krapp, 1992a, S. 305). Dies können Dinge, Lebewesen, Zustände, Veränderungen, Ereignisse und Zusammenhänge sein, womit auch Physikunterricht, dessen Inhalte oder physikspezifische Aktivitäten (z. B. Experimentieren) *Gegenstände* darstellen können.

Das oben definierte Interessenkonstrukt beinhaltet die beiden Komponenten *individuelles* und *situationales Interesse*, die ihren Ursprung in zwei unterschiedlichen Forschungslinien haben (vgl. Krapp, 1992b).

Definition (Individuelles Interesse): „Das individuelle Interesse kann als relativ dauerhaftes, dispositionales Merkmal einer Person verstanden werden, das sich in der Auseinandersetzung mit einem Gegenstandsbereich (z. B. Schulfach) entwickelt und als mehr oder weniger starke Wertschätzung dieses Bereichs zum Ausdruck kommt“ (Schiefele, 2009, S. 163).

Definition (Situationales Interesse): „Das situationale Interesse bezeichnet dagegen den durch äußere Umstände (z. B. einen spannenden Vortrag) hervorgerufenen Zustand des Interessiertseins, der u. a. durch eine erhöhte Aufmerksamkeit und Gefühle der Neugier und Faszination gekennzeichnet ist“ (Schiefele, 2009, S. 164).

Für eine genauere Auseinandersetzung mit dem Konstrukt und dessen Zusammenhang mit motivational günstigen Bedingungen sei an dieser Stelle auf die Literatur verwiesen (Deci & Ryan, 1985, 1993; Krapp, 1992a, 2004, 2005).

Als Rechtfertigung für die intensive Beschäftigung mit dem Thema in diesem Kontext sind zwei wichtige Gründe dafür festzuhalten, die Interessenförderung als ein Ziel schulischen Lernens zu verstehen (Krapp, 1998; Prenzel, Schütte & Walter, 2007). Zum einen gibt es zahlreiche empirische Nachweise dafür, dass „eine auf persönlichen Interessen beruhende Lernmotivation unter bestimmten Voraussetzungen positive Effekte auf die Art und Weise der Lernsteuerung [...], die Erlebensqualität während des Lernens [...] und den kurz- und langfristigen „Output“ des Lernens hat“ (Krapp, 1998, S. 187). Zum Anderen gelten Interessen „als wichtige Komponenten gelungener Bildung“ (ebd., S. 187). Diese eher pädagogisch-bildungspolitische Sichtweise beachtet zudem, dass hohe Kompetenzen in einem Schulfach nur dann wirklich sinnvoll sind, wenn bei den Schülerinnen und Schülern die Bereitschaft (Interesse) besteht, diese weiterzuentwickeln und im späteren Leben (z. B. im Beruf) einzusetzen (Prenzel et al., 2007). Die Förderung des

Interesses an Physik ist damit sowohl normativ gefordert als auch für ein erfolgreiches Lernen zweckmäßig.

Die naturwissenschaftlichen Schulfächer, allen voran *Physik*, sind in Beliebtheitsrankings typischerweise auf den untersten Plätzen. Bei Zwick und Renn (2000) wird *Physik* von 30 %¹ der Schülerinnen und Schüler als unbeliebtes Fach genannt, während nur 10 % es als Lieblingsfach nennen. Bei Fächern wie *Sport* oder *Englisch* ist die Situation umgekehrt. *Sport* wird von 8 % als unbeliebtes Fach und nur von 29 % als Lieblingsfach genannt. Bei *Englisch* ist das Verhältnis sogar noch eindeutiger (4 % zu 17 %). *Mathematik* polarisiert sehr stark, sodass das Verhältnis zwar ausgeglichen, die Anzahl der Nennungen aber relativ hoch ist (20 % zu 21 %). Auch Kessels und Hannover (2004a) bilden eine Rangfolge der *empfundene Selbstnähe* zu zwölf Unterrichtsfächern. *Physik* landet in der Gesamtstichprobe auf den letzten Platz. Bei den Jungen ist die *Selbstnähe* lediglich zu *Musik* und bei den Mädchen nur zu *Geschichte* noch geringer. Interessanterweise ist *Mathematik* hier auf den oberen Rängen platziert (3. Platz gesamt, ♂2. Platz, ♀6. Platz).

Die mangelnde Beliebtheit von *Physik* spiegelt sich zum Teil auch im Interesse wider. Die IPN-Interessenstudie (Hoffmann, Häußler & Lehrke, 1998) untersucht den Einfluss von außerschulischen, Unterrichts- und Persönlichkeitsmerkmalen auf das Fachinteresse und auf das durch den Physikunterricht induzierte Interesse an Physik. Außerdem wird als weitere abhängige Variable das Sachinteresse untersucht, welches bestimmte Kontexte (z. B. Alltag, Beruf und Wissenschaft), bestimmte Inhaltsgebiete (z. B. Optik, Wärme und Radioaktivität) und bestimmte Tätigkeiten (z. B. Beobachten, Experimentieren und Rechnen) berücksichtigt. Die Studie ist als echter Längsschnitt (5. bis 10. Jahrgangsstufe) mit zusätzlich umfangreicherem Querschnitt (9. Jahrgangsstufe) ausgelegt. Zum Fachinteresse kann zusammenfassend festgehalten werden, dass über alle Jahrgangsstufen hinweg die Mädchen ein signifikant geringeres Fachinteresse haben als die Jungen. Während es bei den Mädchen stetig abfällt, kann bei den Jungen ein zunächst zurückgehendes (zwischen Klasse 7 und 8) und zum Ende wieder ansteigendes Interesse beobachtet werden. Im Vergleich zu den Fächern *Chemie*, *Mathematik* und *Biologie* ist das Fachinteresse *Physik* bei den Mädchen in allen Jahrgangsstufen am geringsten. Das trifft auch für den Vergleich mit anderen Fächern zu (*Deutsch*, *Kunst* und *Fremdsprachen*), der zudem deutlich macht, dass das absolute Niveau für *Physik* verhältnismäßig gering ist. Bei den Jungen ist die Situation eine andere. Über alle Jahrgangsstufen hinweg ist das Fachinteresse *Physik* mit *Chemie*, *Mathematik* und *Biologie* zusammen auf einem relativ hohen Niveau und außerdem höher als jenes für *Deutsch*, *Kunst* und *Fremdsprachen*. Die Analysen ergaben zudem u. a. qualitative Unterschiede im Sachinteresse (Interesse-Typen A, B und C, Häußler, Hoffmann, Langeheine, Rost & Sievers, 1996), die jedoch hier nicht genauer betrachtet werden sollen.

Die Ergebnisse der PISA-Studie von 2006 (Organisation für wirtschaftliche Zusammenarbeit und Entwicklung [OECD], 2007a, 2007b) lassen aktuellere Aussagen über

¹höchster Wert unter allen Fächern

das Interesse an Naturwissenschaften in Deutschland zu.² Im Rahmen der Schülerfragebögen wurden mehrere Skalen zur Erfassung der Einstellung zu den Naturwissenschaften eingesetzt. Darunter waren die Skalen *Allgemeines Interesse an Naturwissenschaften* und *Freude an Naturwissenschaften* (siehe auch Tabelle 7.1, S. 104). Das allgemeine Interesse an Naturwissenschaften variiert innerhalb des Fragenkataloges sehr stark. Während 77 % hohes oder durchschnittliches Interesse an der Biologie des Menschen haben, sind es bei Themenbereichen in der Physik „nur“ 56 %. Dies deckt sich mit den Ergebnissen der IPN-Interessenstudie. Dennoch liegt der Index des allgemeinen Interesses an Naturwissenschaften in Deutschland höher als der OECD-Durchschnitt und im obersten Fünftel der OECD-Länder. Beim Index der Freude an Naturwissenschaften liegt Deutschland sehr knapp unter dem OECD-Durchschnitt. Im Allgemeinen macht es 63 % der Schülerinnen und Schülern Spaß, sich mit naturwissenschaftlichen Themen zu befassen und 60 % sind interessiert, Neues in den Naturwissenschaften zu lernen. Dagegen beschäftigen sich nur 38 % gern mit naturwissenschaftlichen Problemen. Bei beiden Indizes gibt es keine empirisch bedeutsamen Geschlechtsunterschiede ($|d| < 0.14$). Dagegen ist der Zusammenhang mit den Leistungstests in beiden Fällen sehr hoch. Ein Anstieg um eine Einheit der genannten Interesse-Indizes bedeutet eine Leistungsverbesserung von ca. 30 Punkten.³ Dieser Leistungsunterschied entspricht nach einer groben Faustformel dem Zuwachs eines Schuljahres (vgl. Prenzel, Carstensen, Frey, Drechsel & Rönnebeck, 2007).

Im aktuellen IQB⁴-Ländervergleich 2012 lag der Schwerpunkt bei den mathematischen und naturwissenschaftlichen Kompetenzen am Ende der Sekundarstufe I. Neben den jeweiligen Kompetenzen wurden im Rahmen der Studie aber auch die motivationalen Schülermerkmale *Selbstkonzept* und *Interesse* erhoben und analysiert (Jansen, Schröders & Stanat, 2013). Auf das physikbezogene Selbstkonzept wird in Abschnitt 5.2 ausführlicher eingegangen. Die Daten erlauben bei dem fachbezogenen Interesse eine Differenzierung der Ausprägungen in den mathematisch-naturwissenschaftlichen Fächern *Mathematik*, *Biologie*, *Chemie* und *Physik*, die mit den Daten der PISA-Studie von 2006 nicht möglich ist. Pro Schulfach gab es vier Aussagen, die jeweils auf einer vierstufigen Ratingskala bewertet wurden. Anhand der Skalenmittelwerte (M) wurde der Anteil der Schülerinnen und Schüler ermittelt, die ein niedriges ($M \leq 2$), ein mittelhohes ($2 < M < 3$) oder ein hohes ($M \geq 3$) Interesse besitzen.

Für das Fach Physik haben 50 % der Schülerinnen und Schüler geringes und nur 22 % hohes Interesse (siehe Abb. 3.1). Die Betrachtung getrennt nach dem Geschlecht macht jedoch deutlich, dass bei den Jungen ein ausgewogenes Verhältnis zwischen den Inter-

²Hier ist zu beachten, dass sich alle hier vorgestellten Ergebnisse dieser Studie auf *Naturwissenschaften* beziehen und ein physikspezifisches Interesse nicht ausdifferenziert werden kann. Die im Rahmen der Leistungstest ebenfalls erhobenen Daten zum themenspezifischen *Lerninteressen* liegen nur teilweise vor.

³Die IRT-skalierten Daten sind so normiert, wie es für internationale Vergleichsstudien üblichen ist. Der OECD-Durchschnitt liegt bei $M = 500$ ($SD = 100$).

⁴Institut zur Qualitätsentwicklung im Bildungswesen

	niedrig	mittel	hoch
Jungen	36%	31%	34%
Mädchen	64%	25%	10%
Gesamt	50%	28%	22%

Abb. 3.1 Geschlechtsbezogene Unterschiede im Interesse an Physik. Die Balken geben die Anteile von Schülerinnen und Schülern mit niedrigem ($M \leq 2$), mittlerem ($2 < M < 3$) und hohem Interesse ($M \geq 3$) an. Die Prozentwerte sind gerundet und weichen dadurch in der Summe von 100 ab. (angelehnt an Jansen et al., 2013, S. 356f)

essenniveaus besteht und bei den Mädchen sich die Situation noch drastischer darstellt. Fast zwei Drittel der Schülerinnen haben nur ein geringes Interesse an Physik und nur eines von zehn Mädchen ist an dem Fach stark interessiert. Der Unterschied zwischen Jungen und Mädchen beim Skalenmittelwert ist signifikant im Bereich eines mittleren Effekts ($d = 0.72$, $p < .05$)⁵.

Der Vergleich mit den anderen mathematisch-naturwissenschaftlichen Fächern zeigt, dass auch hier das Interesse an Physik am geringsten ist (siehe Abb. 3.2). Über beide Geschlechter hinweg ist das Interesse an Mathematik am höchsten, gefolgt von Biologie und Chemie. Bei den Jungen (blaue Säulen) sind die Unterschiede zwischen den Fächern nur gering ($2.45 \leq M \leq 2.71$). Bei den Mädchen (rote Säulen) zeigt sich jedoch, dass sie vor allem an Biologie interessiert sind ($M = 2.56$), mit deutlicher Abweichung von Physik ($M = 1.94$). Auch der Geschlechtsunterschied ist bei keinem anderen Fach so hoch. Bei Biologie ist er empirisch nicht bedeutsam ($|d| < 0.20$) und bei Mathematik und Chemie im Bereich eines kleinen Effekts ($d < 0.5$).

Der positive Zusammenhang zwischen Interesse und Ergebnis im Leistungstest findet sich auch beim IQB-Ländervergleich wieder. Je höher das erreichte Kompetenzniveau in Physik ist, desto höher ist auch der Anteil der hoch interessierten Schülerinnen und Schülern. Allerdings ist dieser Anteil selbst bei den Leistungsstärksten gering. Unter den 25 % der Schülerinnen und Schülern mit dem höchsten Kompetenzniveau haben nur 30 % hohes Fachinteresse, d. h. der größte Anteil von ihnen (70 %) ist nicht an Physik interessiert. Und auch bei den leistungsstärksten 5 % der Schülerinnen und Schüler haben nur die Hälfte ein hohes Interesse für Physik. In dieser Gruppe der Leistungsstärksten aber Nicht-Interessierten sind die Mädchen überrepräsentiert (vgl. Jansen et al., 2013, S. 362).

In allen hier vorgestellten Studien bestätigt sich der von Krapp (1998) dargelegte Zusammenhang zwischen Lernleistung und Interesse (siehe oben), womit das lernpsychologische Motiv der Interessenförderung verstärkt wird. Die kausale Beziehung zwischen

⁵Ein Überblick über die in der vorliegenden Arbeit verwendeten Effektstärken mit der von Cohen (1992) vorgeschlagenen Einteilung in *schwache*, *mittlere* und *starke* Effekte findet sich in Tabelle C.1 im Anhang.

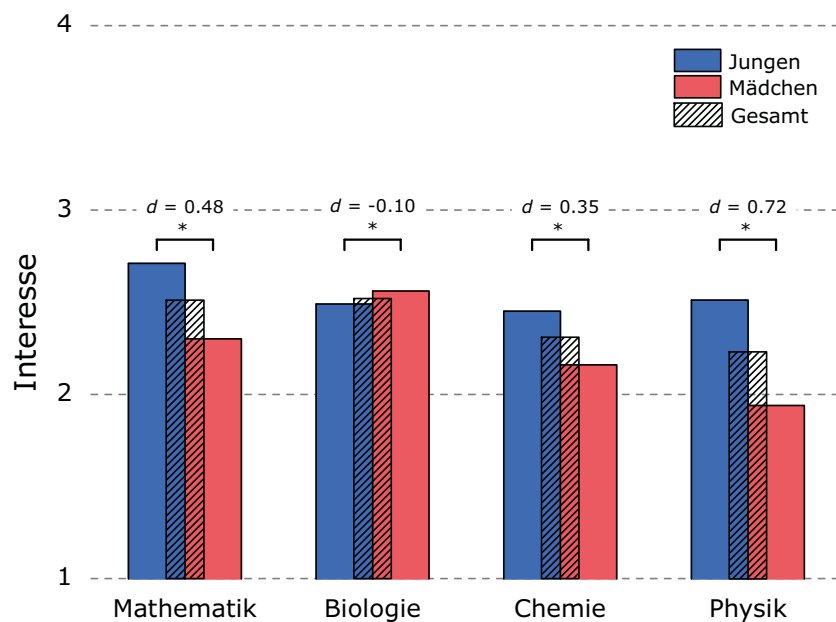


Abb. 3.2 Mittleres und geschlechtsspezifisches Interesse an den mathematisch-naturwissenschaftlichen Fächern (Daten entstammen Jansen et al., 2013, S. 352ff)

Lernleistung und Interesse ist jedoch nicht trivial, da vermutlich von einer wechselseitigen Verstärkung auszugehen ist. Die Ergebnisse des IQB-Ländervergleichs zeigen, dass sich an der Interessenlage im Fach Physik seit den Erhebungen der IPN-Interessenstudie wenig geändert hat. Während die Jungen ein angemessenes und ausgeglichenes Interesse besitzen, welches mit den anderen naturwissenschaftlichen Fächern vergleichbar ist, finden sich nur wenig hoch interessierte Mädchen. Bei Ihnen grenzt sich Physik ganz klar von den anderen Fächern ab. Dies ist insbesondere deswegen ein Problem, weil sehr viele von den leistungsstärksten Schülerinnen, also denjenigen, die potenziell für MINT⁶-Studiengänge und -Berufe infrage kommen, sich aufgrund fehlender Interessen nicht für diese Fachrichtungen entscheiden. Damit wird auch das pädagogisch-bildungspolitische Motiv der Interessensförderung verstärkt, da nicht zuletzt aus volkswirtschaftlichen Gründen noch immer Bedarf an naturwissenschaftlich interessierten Bürgerinnen und Bürger besteht.

3.2 Image von Physik

Eine positive Interessenentwicklung ist nach Kessels und Hannover (2004b) nicht nur von motivational günstigen Bedingungen für z. B. das situationale Interesse abhängig (siehe oben und vgl. Krapp, 1998), sondern auch von *sozial geteilten* Bedeutungen der verschiedenen Interessengegenstände. Sozial geteilte Annahmen über die Merkmale ei-

⁶Akronym für Mathematik, Informatik, Naturwissenschaften und Technik

nes „Gegenstandes“ werden als *Image* bezeichnet. Diese Definition entspricht auch der umgangssprachlichen Verwendung des Begriffes. Das Image von Physikunterricht ist demnach eine subjektive Annahme über die Eigenschaften des Faches, die innerhalb einer Peergroup geteilt wird. Typische Vertreter dieses „Gegenstandes“, die wiederum sozial geteilte Merkmale aufweisen, werden als *Prototypen* bezeichnet. Im Falle eines Schulfaches sind die Prototypen also Personen, die mit dem Fach identifiziert werden, d. h. deren Lehrer oder Schülerinnen und Schüler, die dieses Fach z. B. mögen, nicht mögen oder darin besonders leistungsstark sind. Die sozial geteilten Annahmen über ein Fach übertragen sich auch auf dessen Prototypen und umgekehrt (Kessels, 2008). Im Folgenden soll zunächst kurz ein theoretisches Rahmenkonzept vorgestellt werden, mit dem der Zusammenhang zwischen *Image*, *Prototypen*, *Selbstkonzept* und *Interessenentwicklung* beschrieben werden kann. Anschließend werden die Prototypen und das Image von Physik dargelegt und ein Bezug zu den Assoziationen zu Physik hergestellt.

Das Entstehen von Interesse im Unterricht wird nach Krapp (1998) im Wesentlichen durch die erlebte eigene Kompetenz, Autonomie und soziale Eingebundenheit (*Basic Needs*, vgl. Deci & Ryan, 1985) beeinflusst. Kessels und Hannover (2004b, 2006) betonen jedoch, dass zusätzlich das Image des Interessengegenstandes und das Ansehen der jeweiligen Prototypen von Bedeutung sind. Dabei kommt es vor allem darauf an, inwiefern dieses Image zum eigenen Selbstkonzept⁷ passt und ob man sich selbst mit dem jeweiligen Prototyp identifiziert. Der Hintergrund dafür ist die zentrale Annahme, dass die Interessenentwicklung in der Schule mit der gleichzeitig stattfindenden Identitätsentwicklung im Jugendalter zusammenhängt (Kessels, 2008). Die Entwicklung des eigenen Selbstbildes geschieht zum Teil bewusst und reflektierend. Demnach wäre es zweckmäßig, die in der Peergroup bekannten Eigenschaften von Prototypen zu nutzen, um dem idealen Selbstbild näher zu kommen. Dies kann zum einen geschehen, indem man sich einem bestimmten Fach zuwendet, dessen Image sich gut ins eigene Selbstbild integriert. Zum Anderen kann aber auch durch die Ablehnung eines bestimmten Faches signalisiert werden, wie man sich selbst mit den Prototypen des Faches identifiziert. Schülerinnen und Schüler, die sich von Physik abwenden, bekräftigen so ihre Rolle als typische „Physik-Hasser“ oder differenzieren sich bewusst vom typischen „Physik-Fan“.

Diese Interessenentfaltung als Folge der Identitätsregulierung würde u. a. die Geschlechtsunterschiede im Fach *Physik* erklären. Die Entwicklung des Selbstbildes während der Pubertät, also genau in den Altersstufen, in denen das Interesse der Mädchen an Physik stark nachlässt (siehe Abschnitt 3.1), ist vor allem durch die Bildung der Geschlechterrolle bestimmt (Kessels & Hannover, 2004b). Dies verursacht oftmals sogar ein übertriebenes geschlechtsrollenkonformes Verhalten. Wenn nun Physik ein Image als typisches Jungenfach besitzt, wäre es gerade für Mädchen von Vorteil, ihre „neue“ Rolle als Frau auszuleben, indem sie sich von diesem Fach abwendet (Kessels, 2008).

⁷Die Definition des Konstrukts *Selbstkonzept* erfolgt ausführlich in Abschnitt 5, wo es im Rahmen der *Balanced Identity Theory* als Interpretation von bestimmten Assoziationen angesehen wird.

Hannover und Kessels (2004) gehen zunächst der Frage nach, wie die Prototypen verschiedener Schulfächer von Schülerinnen und Schülern der 8. und 9. Klasse charakterisiert werden. Dazu vergleichen sie die Einschätzungen zu vier Prototypen bezüglich der Eigenschaften *physischer & sozialer Attraktivität, Sozialkompetenz & Integriertheit, Intelligenz & Motivation, Arroganz & Egoismus* und *Kreativität & Emotionalität*. Die vier Prototypen sind Schülerinnen und Schüler, die 1. Naturwissenschaften⁸ am meisten mögen, 2. Naturwissenschaften am wenigsten mögen, 3. Geisteswissenschaften⁹ am meisten mögen und 4. Geisteswissenschaften am wenigsten mögen. Die Ergebnisse zeigen ein deutliches Bild: Schülerinnen und Schüler, die Physik mögen, sind weniger attraktiv (physisch und sozial), isolierter, weniger integriert, weniger kreativ und weniger emotional als diejenigen, die Physik nicht mögen. Dafür werden sie als arroganter und egozentrischer eingeschätzt. Allerdings werden sie auch im Vergleich zu den drei anderen Prototypen als besonders intelligent und motiviert beurteilt. Diese sozial geteilten Annahmen über die Eigenschaften der Prototypen gehen laut Ergebnissen der Studie außerdem damit einher, dass das Selbstbild der Schülerinnen und Schüler den „Physik-Fans“ weniger ähnlich ist, als denjenigen, die Physik nicht mögen. Das lässt vermuten, dass sowohl Mädchen, als auch Jungen, deren ideales Selbst nicht mit den hier beschriebenen Eigenschaften konform ist, eher Schwierigkeiten damit haben werden, sich mit Physik zu identifizieren.

Mit stärkerem Fokus auf geschlechtsbezogene Stereotype vergleicht Kessels (2005) den typischen Jungen und das typische Mädchen, die jeweils Physik oder Musik als Lieblingsfach haben. Das ergibt vier verschiedene Prototypen. Dabei wird vor allem deutlich, dass die Vorliebe für ein Schulfach sehr stark die wahrgenommene Männlichkeit und Weiblichkeit bei der Peergroup mitbestimmt. Am stärksten als männlich oder weiblich eingeschätzt werden jeweils die stereotyp-konformen Prototypen, d. h. der typische Junge, der Physik mag, ist am männlichsten und das typische Mädchen, das Musik mag, ist am weiblichsten (siehe Abb. 3.3). Den zweithöchsten Wert bei dieser Einschätzung erhält jedoch jeweils das andere Geschlecht, dass ebenfalls Musik oder Physik mag. Das bedeutet insbesondere, dass gleich nach dem Jungen, das typische Mädchen, das Physik mag, als männlicher bewertet wird, als der Junge und das Mädchen, die Musik als Lieblingsfach haben.

Des Weiteren deckt die Studie (Kessels, 2005) geschlechtsbezogene Unterschiede der vermuteten Popularität bei den Mitschülern auf, je nachdem, ob ein Mädchen oder ein Junge in Physik gute Leistungen erzielt. So wird von ihnen angenommen, dass Jungen und Mädchen, die in Musik erfolgreich sind, bei den Mädchen beliebter sind als diejenigen, die in Physik leistungsstark sind. Bei den Jungen wird jedoch eine differenzierte Beliebtheit vermutet. Bei ihnen nimmt die Peergroup an, dass sie gute Leistungen der Jungen in Physik genauso wie gute Leistungen der Mädchen in Musik mit Beliebtheit honorieren. Dagegen sind die leistungsstarken Mädchen in Physik bei ihnen ähnlich

⁸gemittelt über die Einschätzungen bezüglich der Fächer *Mathematik* und *Physik*

⁹gemittelt über die Einschätzungen bezüglich der Fächer *Deutsch* und *Englisch*

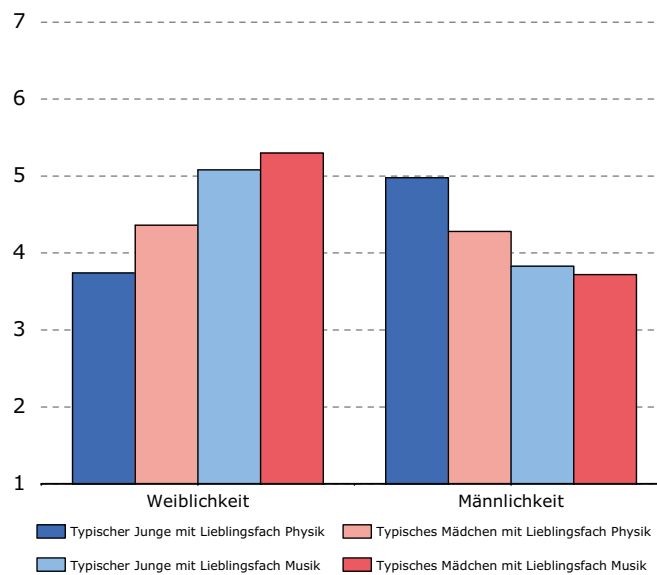


Abb. 3.3 Zugeschriebene Männlichkeit und Weiblichkeit zu männlichen und weiblichen Prototypen, die Musik oder Physik als Lieblingsfach haben (angelehnt an Kessels, 2005, S. 315)

unbeliebt, wie die erfolgreichen Jungen in Musik. In beiden Fällen wäre es für Mädchen vom Nachteil, in Physik durch gute Leistungen aufzufallen.

Es finden sich auch Studien, die sich direkt mit den sozial geteilten Eigenschaften von Schulfächern befassen. Anhand derer Ergebnisse lassen sich Aussagen über das Image von Physik treffen, ohne von den Eigenschaften der Prototypen darauf schließen zu müssen. Watson, McEwen und Dawson (1994) stellen mithilfe von Repertory Grids Rangfolgen der Schulfächer für verschiedene Eigenschaftsdimensionen auf. Die Dimensionen *Difficulty* und *Freedom* fallen im besonderen Maße auf, da bei ihnen die mathematisch-naturwissenschaftlichen Fächer jeweils die vier ersten bzw. letzten Plätze der Rangfolge einnehmen. Anhand dieser Eigenschaften lassen sich die MINT-Fächer also klar von den sprachlich-geisteswissenschaftlichen Fächern unterscheiden. Physik ist jeweils das Fach mit der zweitstärksten Ausprägung. Zum einen setzt es logisches Denken voraus, erfordert es ein natürliches Talent, hohe Problemlösefähigkeit sowie das Verständnis abstrakter Prinzipien und ihm liegen schwierige und verwirrende Ideen zugrunde (*Difficulty*). Zum anderen ist Physik ein Fach, welches keine offenen Antworten zulässt, keine Kreativität fordert, wenig Gelegenheit zur Selbstdarstellung bietet und selten die Emotionen und Gefühle der Schülerinnen und Schüler anspricht (*Freedom*).

Neben zusätzlichen Hinweisen dafür, dass Physik als schwierig und fremdbestimmt wahrgenommen wird, liefern Hannover und Kessels (2002) auch Ergebnisse von weitergehenden Untersuchungen, die die Geschlechtskonnotation verschiedener Fächer aufdecken. Demnach wird Physik (nach Sport) am häufigsten im Vergleich zu den anderen Fächern als „Jungenfach“ bezeichnet, während Musik und Kunst ganz klar weiblich

konnotiert sind. Diese Geschlechtskonnotation von Physik ist bei den Mädchen stärker ausgeprägt als bei den Jungen.

Die drei Image-Dimensionen *Schwierigkeit*, *Männlichkeit* und *Fremdbestimmtheit* werden von Kessels, Rau und Hannover (2006) für Physik auch auf der Ebene von automatischen Stereotypisierungen mithilfe eines Tests zu Messung impliziter Assoziationen (IAT)¹⁰ untersucht. Dem liegt die Annahme zugrunde, dass die Zuschreibung der Merkmale von Unterrichtsfächern bzw. die Aktivierung des Images auch unbewusst und automatisch passiert, z. B. bei einer Konfrontation mit einem Prototyp. Sie können zunächst zeigen, dass das aus anderen Studien bekannte Image von Physik auch in Form von Assoziationsmaßen messbar ist. Die ermittelten Effekte entsprechen demnach gedankliche Verknüpfungen zwischen *Physik* und *schwierig*, *männlich* sowie *fremdbestimmt*. Bei dem auf dieser Ebene der Assoziationen gemessenen Image von Physik finden sich die gleichen geschlechtsbezogenen Unterschiede wieder, wie bei den Studien mit expliziten Äußerungen (z. B. Hannover & Kessels, 2002; Kessels, 2005). Die Mädchen haben stärkere Assoziationen zwischen *Physik* und *schwierig* sowie *männlich* als die Jungen, womit das Fach bei den Mädchen ein negativeres implizites Image hat.

Das implizite Image von Physik erweist sich bei Kessels et al. (2006) (Studie 1) als Prädiktor für explizite Einstellungen zum Fach. Bei der Berücksichtigung aller Image-Dimensionen (*Schwierigkeit*, *Männlichkeit* und *Fremdbestimmtheit*) werden 39 % der Varianz bei den expliziten Einstellungen aufgeklärt und damit mehr, als wenn z. B. nur die Dimension *Männlichkeit* in Betracht gezogen wird (25 %, wie es bei vergleichbaren Studien geschieht (z. B. Nosek, Banaji & Greenwald, 2002, siehe Abschnitt 3.3)). Die Autorinnen vermuten darin, dass das Image von Physik nicht nur von der Geschlechtskonnotation beeinflusst wird, den Grund dafür, dass auch bei den Jungen andere Fächer meist beliebter sind (siehe Abschnitt 3.1).

Kessels et al. (2006, Studie 2) zeigen außerdem, dass das Image von und implizite Assoziationen zu Physik auch verändert werden können. Schülerinnen und Schüler, die in einer kurzen Intervention einen Text aus „Die Struktur wissenschaftlicher Revolutionen“ von Thomas S. Kuhn zu lesen bekamen, sollten im Anschluss daran in einem Satz die Kernaussage des Textes wiedergeben und schließlich zwei implizite Assoziations-tests durchführen. Der erste Test maß die Assoziationen von *Physik* zu *fremdbestimmt* und der zweite Test war ein Einstellungs-IAT, der die Assoziation zu wertenden Adjektiven wie *gut* oder *schlecht* misst. Als Kontrollgruppe dienen Schülerinnen und Schüler, die ein Physik-Lehrbuchtext (Sek. II) über verschiedene Formen von Linsen lesen und Fragen über die äußere Erscheinung des Textes beantworten sollten. Die Kuhn-Gruppe hat schwächere negative Assoziationen zu Physik ($t(15) = -2.41, p < .05, d = 1.0$) und schwächere Assoziationen zu *fremdbestimmt* ($t(19) = -1.93, p = .07, d = 0.84$) als die Lehrbuch-Gruppe (siehe Abb. 3.4).¹¹

¹⁰Das Verfahren wird in Kapitel 4 ausführlich vorgestellt.

¹¹Die Unterschiede zwischen beiden Gruppen treten nur bei Schülerinnen und Schülern auf, die Physik zum Zeitpunkt der Studie noch in der Kursphase besuchen. Bei denjenigen, die das Fach bereits so

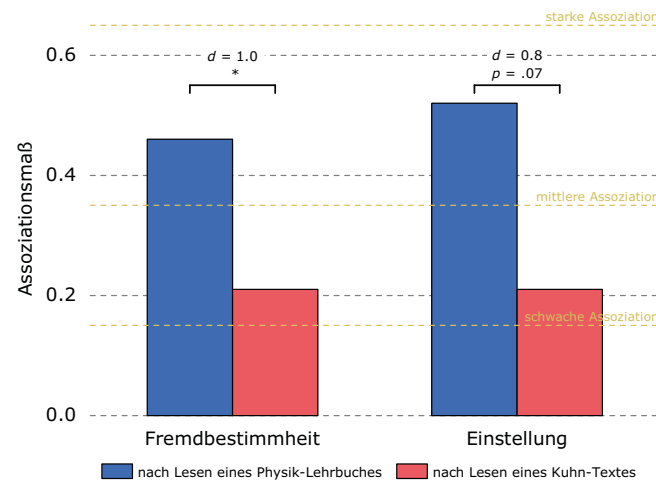


Abb. 3.4 Unterschiedliche Assoziationen zu Physik in zwei verschiedenen Gruppen. Ein positives Assoziationsmaß bedeutet eine Assoziation von *Physik* zu *fremdbestimmt* (links) bzw. *schlecht* (rechts) (Daten entstammen Kessels et al., 2006, S. 772f).

In einer weiteren Studie (genaue statistische Kennwerte liegen nicht vor) berichten Kessels und Hannover (2006) von einer ähnlichen Manipulation, in der die implizite Geschlechtskonnotation abgeschwächt werden konnte. In einer Gruppe lasen die Probanden einen Text über eine fiktive Physikerin und die Kontrollgruppe las den gleichen Text, nur mit einem gegen einen männlichen Physiker ausgetauschten Rollenmodell. In einer dritten Gruppe wurde ein geschlechtsneutraler Text gelesen. Die Assoziation von *Physik* zu *männlich* ist in der Gruppe mit weiblichem Rollenmodell schwächer als in den beiden anderen Gruppen. Das bedeutet, dass der erwartungskonforme männliche Physiker die Geschlechtskonnotation nicht beeinflusst, während eine weibliche Repräsentantin eine schwächende Wirkung darauf haben kann.

Zusammenfassend kann also festgehalten werden, dass das Image des Physik-Prototyps (gute Leistungen und Vorliebe zum Fach) für den größten Teil der Schülerinnen und Schüler keine hohe Anziehungskraft hat. Er gilt als unbeliebt, isoliert und unattraktiv, auch wenn er als intelligent eingeschätzt wird. Dieses schlechte Image scheint für ein Mädchen verhängnisvoller zu sein, da dem Prototyp zusätzlich eher Männlichkeit zugeschrieben wird. Die Konsequenz daraus zeigt sich darin, dass der überwiegende Anteil der Schülerinnen und Schüler wenig Ähnlichkeit zwischen ihrem Selbstbild und dem Physik-Prototyp sehen. Beim Image des Schulfaches Physik lassen sich drei Faktoren identifizieren, anhand deren es sich von sprachlich-geisteswissenschaftlichen Fächern unterscheiden lässt. Entsprechend den Eigenschaften des Prototyps gilt auch das Fach als ein typisches Jungenfach, als sehr anspruchsvoll und fremdbestimmt. Die Fremdbestimmtheit bedeutet im weitesten Sinne auch der Mangel an Möglichkeiten der Selbst-

früh wie möglich ausgewählt haben, sind die Assoziationen in beiden Gruppen nicht signifikant unterschiedlich.

darstellung, an Kreativität und an Berücksichtigung der Emotionen der Schülerinnen und Schüler.

3.3 Assoziationen zu Mathematik und Naturwissenschaften

Neben den Studien zum Image von Physik und den fachbezogenen Assoziationen gibt es auch Untersuchungen zu den Assoziationen bezüglich Mathematik und Naturwissenschaften im Allgemeinen. Falls der angenommene Zusammenhang zwischen den sozial geteilten Annahmen über ein Schulfach (Fachimage) und der Interessenentwicklung besteht (Kessels & Hannover, 2004b) und falls die Messung des Images durch Messung von Assoziationen auch in anderen Fächern angewendet werden kann, so müssten die Assoziationen zu Mathematik und Naturwissenschaften mit der oben dargestellten Interessenlage in diesen Fachbereichen korrespondieren.

Nosek et al. (2002) messen die Assoziationen von Studierenden ($N = 83$) zwischen Mathematik und bewertenden Adjektiven (Einstellung zu Mathematik), dem Selbst der Probanden (mathematikbezogenes Selbstkonzept) sowie dem Geschlecht (Geschlechtsstereotyp).¹² Außerdem werden die Einstellung zu Naturwissenschaften und die Assoziation zwischen dem Selbst und dem Geschlecht gemessen (Geschlechtsidentität). Beide Fachbereiche werden überwiegend negativ bewertet, jedoch von den Frauen sehr viel stärker als von den Männern ($t(77) = 4.24$, $p = 10^{-5}$, $d = 0.97$). Im Gegensatz dazu wird Mathematik über beide Geschlechter hinweg eher mit *männlich* assoziiert. Außerdem gibt es starke Zusammenhänge zwischen der eigenen Geschlechtsidentität und der Identifikation mit Mathematik, passend zu der jeweiligen Geschlechtsstereotypisierung des Fachs. Dies ist von vornherein nicht selbstverständlich, da *Wissen über Stereotypisierungen* und eigene *Vorlieben* als unabhängige Konstrukte angenommen werden. Das heißt, je stärker man ein Schulfach oder eine Wissenschaft mit einer Gruppe assoziiert, mit der man sich selbst identifiziert (in diesem Fall die Gruppe *Männer* oder *Frauen*), desto stärker identifiziert man sich auch selbst mit dem Fach bzw. der Wissenschaft. Es gilt jedoch auch die Umkehrung. Diejenigen, die Mathematik mit der Gruppe assoziieren, mit der sie sich selbst nicht identifizieren, haben auch nur schwache Assoziationen zwischen dem eigenen Selbst und Mathematik. Ein bemerkenswertes Ergebnis der Studie ist außerdem der Zusammenhang zwischen der Mathematik/Geschlechts-Stereotypisierung und dem Abschneiden in einem Mathematik-Leistungstest. Obwohl es im Testergebnis in dieser Stichprobe keine signifikanten Unterschiede zwischen Männern und Frauen gibt, bedeuten hohe Assoziationen von *Mathematik* mit *männlich* für Männer ein besseres Abschneiden im Test ($r = .51$, $p = .0007$) und für Frauen ein schlechteres ($r = -.16$, $p = .30$). Das ist ein Hinweis dafür, dass die Stärke der Assoziation zwischen einem Schulfach und dem Geschlecht mit der Leistung in dem Fach zusammenhängt, je nachdem, zu welchem Geschlecht man gehört. Sowohl die geschlechtsbezogenen Unterschie-

¹²Zur Interpretation der Assoziationen als das jeweilige Konstrukt in Klammern sei auf Abschnitt 5 verwiesen

de bei den fachbezogenen Selbstkonzepten, bei den Einstellungen zu den Fächern, deren insgesamt negativen Tendenzen als auch die komplexen Zusammenhänge zum Ergebnis im Leistungstest stehen im Einklang mit den Ergebnissen der aktuellen Interessenstudien (OECD, 2007a; Jansen et al., 2013, siehe Abschnitt 3.1) und unterstützen damit die oben getroffenen Annahmen.

Cvencek, Meltzoff und Greenwald (2011) untersuchen bei Grundschulkindern (Alter: 6 bis 11 Jahre), ob die geschlechtsstereotypischen Assoziationen, die von älteren Schülerinnen und Schülern sowie von Erwachsenen bekannt sind, bereits zu Beginn der schulischen Laufbahn auftreten. Dabei werden die Assoziationen zwischen Mathematik und dem Geschlecht sowie dem Selbst implizit und explizit gemessen. Die Jungen verbinden Mathematik gedanklich stärker mit dem eigenen Geschlecht als die Mädchen ($t(220) = 6.46, p < .001, r = .40$)¹³. Auch die Assoziation zu dem Selbst, also das mathematikbezogene Selbstkonzept, ist bei den Jungen stärker als bei den Mädchen, allerdings ist der Unterschied hier geringer ($t(220) = 2.63, p < .01, r = .17$, kleiner Effekt, siehe Fußnote 13). Damit zeigen sich schon bei Schulanfängern sowohl auf impliziter als auch auf expliziter Ebene Stereotypisierungen, die sich zumindest augenscheinlich auch auf das Selbstkonzept auswirken.

Wenn (implizite) Assoziationen zu einem Schulfach das Interesse und das fachbezogene Selbstkonzept beeinflussen, dann wäre es auch naheliegend, dass wie beim Interesse und beim Selbstkonzept Zusammenhänge zur schulischen Leistung bestehen. Nosek et al. (2009) können zeigen, dass Unterschiede zwischen Männern und Frauen bei den geschlechtsstereotypischen Assoziationen die Geschlechtsunterschiede in einem Leistungstest vorhersagen können. Bei einer sehr großen internationalen Stichprobe ($N > 500000$) führen sie implizite und explizite Messungen zu den Assoziationsstärken zwischen *science* und *male* durch. Aufgrund der großen Teilnehmerzahl können sie für jedes der 34 teilnehmenden Länder eine mittlere „nationale Assoziationsstärke“ bestimmen. In allen Nationen werden Naturwissenschaften eher mit männlich als mit weiblich assoziiert (>70 % aller Teilnehmer), die Ausprägung variiert allerdings zwischen schwacher und sehr starker Assoziation¹⁴. Als abhängige Variable wird der Geschlechtsunterschied in den Leistungstests (Naturwissenschaften und Mathematik) von TIMSS 2003 herangezogen. Auch hier wird für jede Nation eine Differenz aus der errichteten Punktzahl der männlichen und weiblichen Teilnehmer bestimmt. Die Regressionsanalysen zeigen, dass mit Zunahme der geschlechtsstereotypischen Assoziation um eine Standardabweichung der Leistungsunterschied zwischen Männern und Frauen um 0.56 Standardabweichungen wächst. Von Bedeutung ist allerdings an dieser Stelle, dass nur die impliziten Assoziationen zu diesem starken Zusammenhang mit den Leistungsunterschieden beitragen. Der angenommene lineare Zusammenhang würde bis zu 24 % der Varianz bei

¹³Die Effektstärke (Pearsons r) wurde vom Autor mithilfe der Kennwerte der t -Statistik berechnet: $r = \sqrt{\frac{t^2}{t^2 + df}}$. Sie entspricht einem mittleren Effekt, siehe Tabelle C.1 im Anhang.

¹⁴ $0.28 \leq D \leq 0.65$; Details zum D -Wert (Assoziationsmaß) siehe Abschnitt 4.1.2

den Punktedifferenzen aufklären. Die expliziten Assoziationen leisten nur eine zusätzliche Varianzaufklärung von 2 %.

Die hier aufgeführten Studien zu den Assoziationen zu Mathematik und Naturwissenschaften zeigen, dass zumindest in den berücksichtigten Dimensionen (Geschlecht, Selbstkonzept) eine große Ähnlichkeit zum Image von Physik besteht. In allen Studien wurden Tests eingesetzt, welche die Assoziationen sowohl implizit als auch explizit erfassen. Die so gewonnenen Ergebnisse sind erwartungskonform zu der Annahme, dass die Assoziationen zu einem Schulfach bzw. das Image einen Einfluss auf die Interessenentwicklung haben, da die Struktur Ähnlichkeiten mit der oben dargestellten Interessenlage aufweist und außerdem vergleichbare Zusammenhänge zur Leistung im Fach bestehen. Letzteres allein würde jedoch schon eine angestrebte Einflussnahme auf die Assoziationen bzw. das Image rechtfertigen.

3.4 Bedeutung für die vorliegende Arbeit

Interesse an den Inhalten des Physikunterrichts gilt als Bildungsziel und Voraussetzung für erfolgreiches Lernen. Aus diesen Gründen ist es auch eine Zielsetzung von phänomenbasiertem Physikunterricht. Die aktuelle Interessenlage zur Physik macht deutlich, dass dieses Anliegen nicht an Bedeutung verloren hat. Als Voraussetzung für eine positive Interessenentwicklung gilt ein zum Selbstkonzept passendes Fachimage. Das Image kann auch als Gesamtheit der impliziten Assoziationen zum Fach verstanden werden. Sie können durch äußere Faktoren, wie dem Physiklehrer (Prototyp) oder dem Fachraum aktiviert werden. Diese impliziten Assoziationen lassen sich durch geringfügige Variationen des Unterrichts verändern. Damit ist eine Überprüfung auf Ebene der impliziten Assoziationen möglich, inwiefern phänomenbasierter Physikunterricht es schafft, das Image von Physikunterricht zu verbessern, was dem Ziel der Interessenförderung dienlich wäre.

Das fachbezogene Selbstkonzept steht in einem engen wechselseitigen Zusammenhang zum Interesse. Die Merkmale PbPUs, die unter dem Aspekt der Mediation zusammengefasst sind, unterstützen ein positives physikbezogenes Selbstkonzept, womit ein günstiger Einfluss des Unterrichts auf das Interesse erklärbar wäre. Dementsprechend hat auch die Steigerung der gefühls- und wertbezogenen Bedeutsamkeit des Lerngegenstandes durch die Aspekte der Subjektivität, Mediation und Affektivität eine Wirkung auf das Physikinteresse (vgl. Schiefele, 2009). Als Beispiel sei hier nur der Raum für emotionale Erlebnisse im PbPU genannt, der die Charakterisierung des Physik-Prototyps (siehe Abschnitt 3.2) verändern kann.

Das Fachimage in Form von impliziten Assoziationen, das Interesse und das fachbezogene Selbstkonzept sind drei wechselseitig wirkende Konstrukte. Sie stellen im Rahmen

der vorliegenden Arbeit die abhängigen Variablen dar, auf welche die Auswirkung Pb-
PUs untersucht werden soll.

Messung von impliziten Assoziationen

Für Aussagen über Assoziationen jeglicher Art, wie sie in den in Kapitel 3 zitierten Studien und schließlich auch in der vorliegenden Arbeit getan werden, ist es unerlässlich, Messverfahren einzusetzen, die gezielt für die Erfassung von Assoziationen entwickelt wurden. Die Sozial-, Kognitions- und Persönlichkeitspsychologie stellen geeignete Methoden bereit, die in der Lage sind, Einstellungen, Stereotypen, Selbstkonzept, Selbstwert und eben Assoziationen zu messen. An dieser Stelle sollen die für die vorliegende Arbeit relevanten Verfahren und Instrumente vorgestellt werden. Dafür sind zunächst *implizite* und *explizite* Methoden zu unterscheiden.

Zu den expliziten Methoden zählen alle diejenigen Messungen, die mehr oder weniger direkt eine Selbstbeurteilung der Probanden erheben. Hofmann, Gawronski, Gschwendner, Le und Schmitt (2005) schlagen eine Kategorisierung in Skalen (z. B. Modern Racism Scale oder Rosenberg Self-Esteem Scale), semantische Differentiale, Gefühlsthermometer, Adjektivbewertungen und Single-Item Messungen vor.

Zu den impliziten Methoden gehören z. B. Priming-Prozeduren, Implicit Association Tests (IATs), Single Target IATs (ST-IATs), Go/No-Go Association Tasks (GNATs), Implicit Association Procedures (IAPs), Wort-Vervollständigungs-Aufgaben und weitere Verfahren, die auf Messungen von Gehirnströmen, Gesichtsmuskelaktivitäten oder kardiovaskuläre Aktivitäten beruhen. Einen detaillierten Überblick mit Beispielstudien aus der Stereotypenforschung geben Fazio und Olson (2003). Laut den Autoren teilen sich die impliziten Methoden die Eigenschaft, das interessierende Konstrukt ohne direkte Befragungen der Probanden zu messen. Oftmals sind sich diese nicht mal darüber bewusst, dass Assoziationen oder Stereotypisierungen bei ihnen gemessen werden, bzw. welche das sind. Nach Ansicht der Autoren haben die impliziten Methoden das Hauptziel, das Problem der sozialen Erwünschtheit zu umgehen, was in der Stereotypenforschung in sensiblen Bereichen (z. B. Rassismus, Vorurteilen gegenüber Geschlechtern usw.) auch

überwiegend zutrifft. Kessels et al. (2006) nennen aber einen noch darüber hinausgehenden Grund. Es gibt Belege der Kognitionsforschung dafür, dass Informationsverarbeitungen, deren Bewertung und folgende Entscheidungen auch automatisch, unbewusst, also implizit stattfinden können (vgl. Fazio, 2001). Das bedeutet insbesondere, dass eine Stereotypisierung z. B. zum Fach *Physik* auch automatisch aktiviert werden kann, wenn es nur einen äußeren Reiz gibt, wie die Konfrontation mit typischen Vertretern des Faches (Prototypen) oder auch das Betreten des Fachraumes. Somit ist es also sinnvoll, die Stereotypisierungen oder Assoziationen auch auf der unbewussten Ebene zu messen, auf der sie stattfinden, was die Verwendung von impliziten Methoden erforderlich macht.

4.1 Implicit Association Test

Zu den am häufigsten eingesetzten und somit sehr erprobten Verfahren gehört der Implizite Assoziationen Test (IAT). Seit seiner Einführung durch Greenwald, McGhee und Schwartz (1998) fand der Test in unterschiedlichsten Gebieten mit zunehmender Häufigkeit Verwendung. Der grundlegende Zeitschriftenartikel (ebd.) gehört laut Science-Watch zu den vier bedeutendsten Veröffentlichungen im Bereich *Psychiatry & Psychology* (vgl. Thomson Reuters, 2009) und wurde bis dato laut Google Scholar 4775 mal zitiert (Google Inc., 2013). Seinen ursprünglichen Einsatzzweck fand der IAT in dem Bereich der Sozialpsychologie zur Messung von individuellen Unterschieden in der impliziten Sozial-Kognition. Da dies aber durch die Messung von impliziten Assoziationen geschieht, ist er überall dort einsetzbar, wo mithilfe einer Assoziation zwischen zwei Konzepten eine für den jeweiligen Einsatzzweck relevante Aussage getroffen werden kann. Das Maß der Assoziation zwischen dem *Selbst* und *positiv* kann z. B. als Selbstwert interpretiert werden, was den Test für die Persönlichkeitspsychologie interessant macht. Mittlerweile wird der IAT in fast allen Bereichen der Psychologie eingesetzt: die klinische Psychologie, Konsumentenpsychologie, Gesundheitspsychologie, Entwicklungspsychologie, Altersforschung und die Neuropsychologie (vgl. die Aufzählung mit Beispielen von Gawronski & Conrey, 2004).

4.1.1 Ablauf eines IATs

Bei dem IAT handelt es sich im Wesentlichen um eine computerbasierte Sortieraufgabe, in der die Reaktionszeit die Messgröße ist. Dem Test liegt die Annahme zu Grunde, dass eine gedankliche Verknüpfung zweier Konzepte (Assoziation) Einfluss darauf hat, wie schnell und fehlerfrei diese Aufgabe bewältigt wird. Je nachdem, ob die beiden Konzepte dieselbe oder unterschiedliche Sortieroptionen sind, fällt die Aufgabe leichter oder schwieriger.

In der Mitte des Bildschirms erscheinen in zufälliger Reihenfolge Wörter oder Bilder (*Stimuli*), die per Tastendruck zwei Kategorien zugeordnet werden müssen. Diese Zuordnung ist immer eindeutig und die Zugehörigkeit der Stimuli zu den Kategorien ist

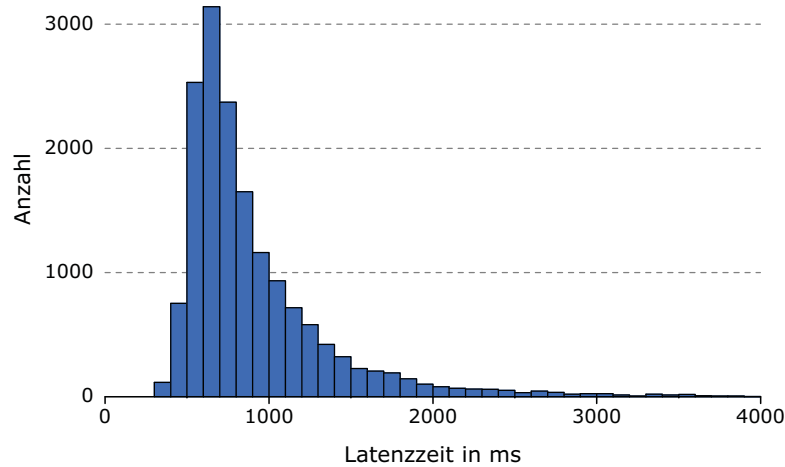


Abb. 4.1 Typische Verteilung der Latenzzeiten eines IATs. Die Daten entstammen aus den beiden kompatiblen Blöcken (Block 3 und 4) einer Stichprobe der Größe $N = 221$.

von vornherein festgelegt. Wählt der Proband falsch, erhält er eine Fehlermeldung (ein rotes **X** in der Mitte des Bildschirms) und er muss korrigieren. Die zur Wahl stehenden Kategorien befinden sich immer rechts und links am oberen Bildschirmrand. Für die Zuordnung nach links muss die Taste *W* und für die Zuordnung nach rechts muss die Taste *P* auf der Tastatur gedrückt werden. Die Probanden werden darum gebeten, die Aufgabe so schnell wie möglich und mit so wenigen Fehlern wie möglich zu bearbeiten. Bei jedem Stimulus wird die Reaktionszeit zwischen Erscheinen des Reizes und korrekter Zuordnung gemessen (*Latenzzeit*). Die Genauigkeit der Messung liegt bei 1 ms und der typische Messwert liegt zwischen 300 ms und 1500 ms (siehe Abb. 4.1). Eine solche Abfolge (Erscheinen des Stimulus – Tastendruck – evtl. Fehlermeldung – Messung der Latenzzeit) wird als *Trial* bezeichnet.

Der bis hier beschriebene Ablauf stellt das allgemeine Schema eines IATs dar. Für die Schlussfolgerung auf eine Assoziation zwischen zwei Kategorien werden diese miteinander kombiniert. Man unterscheidet hierbei immer das im Zentrum stehende Konstrukt (*Ziel*) und die Eigenschaft (*Attribut*) zu der die Stärke der Assoziation des Ziels gemessen werden soll. Im Folgenden soll als Beispiel die Assoziation zwischen *Naturwissenschaften* und dem *Geschlecht* gemessen werden. Der IAT benötigt sowohl für das Ziel, als auch für das Attribut immer zwei sich sinnvoll, gegenseitig ausschließenden Kategorien, die idealerweise aus derselben Domäne stammen (siehe Lane, Banaji, Nosek & Greenwald, 2007, S. 86): „[A] comparison category should be a sensible, mutually exclusive category that is ideally from the same domain (e. g., choosing *humanities* as the companion target category for the academic domain *science*)....“

Eine Zielkategorie steht mit *Naturwissenschaften* fest. Laut den Autoren (siehe Zitat oben) wäre eine angemessene Vergleichskategorie *Geisteswissenschaften*. Für die

4 Messung von impliziten Assoziationen















Block	Trials	Funktion		
1	20	Übung		
2	20	Übung		
3	20	Übung		
4	40	Test		
5	20	Übung		
6	20	Übung		
7	40	Test		

Abb. 4.2 Ablauf eines IATs mit sieben Blöcken. Auf der rechten Seite sind für jeden Block zwei beispielhafte Trials dargestellt. Bei den Blöcken 3 und 4 bzw. 6 und 7 sind die Paarungen auf dem Bildschirm identisch.

Attributdimension *Geschlecht* wären das die beiden Attributkategorien *männlich* und *weiblich*.¹

Der Standard-IAT lässt sich sinnvoll in sieben Blöcke unterteilen.² Der erste Block ist ein *Übungsblock*, in dem die Zuordnung der Stimuli zu den Zielkategorien trainiert werden soll. Der Proband macht sich mit der jetzt immer wiederkehrenden Aufgabe vertraut und lernt die Stimuli kennen, die zur Zieldimension gehören. Dies können in unserem Beispiel Wörter wie *Prosa*, *Philosophie*, *Experiment* und *Atom* sein oder Bilder von typischen Gegenständen, die eindeutig einer der beiden Kategorien zugehörig sind. Da dies nie hundertprozentig der Fall ist³, wird in diesem Block so auch die Zugehörigkeit der Stimuli zu den Kategorien geübt. Am oberen Bildschirmrand stehen die beiden Zielkategorien, zu denen die Stimuli per Tastendruck nach links oder rechts sortiert werden müssen (siehe Abb. 4.2).

Nach 20 Trials endet der erste Block und es folgt mit ebenfalls 20 Trials der zweite Block. Nun erscheinen die beiden Attributkategorien *männlich* und *weiblich* am oberen Bildschirmrand. Dieser Block erfüllt dieselbe Funktion wie der erste, nur dass jetzt die Zuordnung der zu den Attributkategorien gehörigen Stimuli, wie z. B. *Onkel*, *Nichte* oder Namen und Bilder von Männern und Frauen, geübt wird.

Im dritten Block werden Ziel- und Attributdimension kombiniert. Am oberen Bildschirmrand stehen jetzt auf jeder Seite eine Ziel- und eine Attributkategorie (siehe Abb. 4.2, Zeile 3 und 4). Bei vielen Assoziationen, die mit dem IAT gemessen werden, gibt es Vorabannahmen darüber, ob das Ziel eher mit dem einen oder eher mit dem anderen Attribut assoziiert wird. Im hier verwendeten Beispiel kann man für die meisten Zielpopulationen davon ausgehen, dass *Naturwissenschaften* eher mit *männlich* als mit *weiblich* assoziiert werden. Daher wird in der Literatur zum IAT sehr häufig die Bezeichnung (stereotyp-)kompatible/r Paarung/Block verwendet, wenn (auf das Beispiel angewendet) *Naturwissenschaften* und *männlich* gemeinsam auf einer Seite stehen.⁴ Es erscheinen nun alle Stimuli, die bereits aus dem ersten und zweiten Block bekannt sind. Der Proband muss entscheiden, ob ein Stimulus zu *Naturwissenschaften* oder *männlich* gehört und es im positiven Fall auf die Seite sortieren, auf der die beiden Kategorien gemeinsam stehen. Zur Unterstützung sind die Stimuli der Zielkategorien in einer anderen Farbe als die der Attributkategorien (wie auch deren Bezeichnungen auf dem Bildschirm, siehe Abb. 4.2). Die Entscheidung, ob ein Stimulus zum Ziel oder Attribut gehört, ist hier nicht von Interesse und kann damit dem Probanden so weit wie möglich abgenommen

¹Alternativ wären auch *Mann* und *Frau* denkbar. Die Attributkategorien müssen keine Adjektive, sondern können jedes denkbare Konzept wie Personengruppen, Eigenschaften oder Bewertungen sein (siehe 5).

²Sriram und Greenwald (2009) stellen als eine kürzere Version des IATs den Brief Implicit Association Test (BIAT) vor, der aber bei weitem nicht so etabliert und so gut verstanden wie der IAT ist.

³Es wäre z. B. denkbar, dass ein Proband das Wort *Philosophie* zu *Naturwissenschaften* sortiert, weil viele Naturwissenschaftler auch Philosophen waren.

⁴Die Entscheidung, ob bei einem Probanden im dritten Block eine kompatible oder eine inkompatible Paarung auftritt, wird randomisiert getroffen. Im weiteren Verlauf wird immer von einem kompatiblen dritten Block ausgegangen.

werden. Außerdem kann so einer Konfundierung von Ziel und Attribut entgegengewirkt werden. Ungünstige Stimuli wären z. B. *Albert Einstein* oder *Lise Meitner*. Bei ihnen wäre im kombinierten Block nicht klar, ob sie zur Ziel- oder zur Attributdimension gehören, auch wenn sie sich innerhalb dieser Dimensionen den Kategorien klar zuordnen lassen. Die farbliche Markierung hilft in solchen Fällen, auch wenn von der Benutzung konfundierender Stimuli abzuraten ist (siehe Bluemke & Frieze, 2006, mehr Details in Abschnitt 4.1.3). Die ersten 20 Trials dienen wieder der Übung, da sich die Aufgabe an die Probanden von den ersten Blöcken unterscheidet und komplexer geworden ist. In einer kurzen Zwischenmeldung werden die Probanden darum gebeten, die von eben bekannte Aufgabe ein weiteres Mal durchzuführen. Außerdem werden sie noch einmal daran erinnert, die Sortierung so schnell wie möglich und mit so wenigen Fehlern wie möglich durchzuführen. Im Anschluss daran beginnt der vierte und eigentliche Testblock. Er ist doppelt so lang wie der dritte, jedoch die Aufgabenstellung, die Stimuli und die Kombination der Kategorien bleiben unverändert.

Der fünfte Block gleicht in der Länge und in der Funktion dem ersten, jedoch mit einem wichtigen Unterschied. Die Seiten der beiden Zielkategorien sind jetzt vertauscht. Damit sich die Probanden an die neue Anordnung gewöhnen können, führen sie wieder 20 Trials durch.

Im sechsten Block geschieht der für das dem Test zugrundeliegende Paradigma entscheidende Schritt. Wieder werden (wie im dritten Block) Ziel- und Attributkategorien kombiniert. Die Zielkategorie bleibt in ihrer neuen Anordnung aus dem fünften Block. Die Attributkategorien erscheinen jedoch in ihrer ursprünglichen Anordnung. Somit werden aus den kompatiblen Paarungen aus dem dritten Block inkompatible Paarungen. Nun muss bei jedem Stimulus entschieden werden, ob es zu *Naturwissenschaften* oder *weiblich* gehört, um auf die gemeinsame Seite sortiert zu werden. Analog zum dritten/vierten Block üben die Probanden zunächst die Zuordnung in der neuen Konstellation (20 Trials) und führen dann den Testblock durch (siebenter Block, 40 Trials).

Die oben skizzierte grundlegende Annahme des IATs besteht nun darin, dass die Sortieraufgabe dem Probanden in den für ihn stereotyp-kompatiblen Blöcken leichter fällt, als in den stereotyp-inkompatiblen Blöcken.

Wenn der Proband z. B. *Naturwissenschaften* mit *männlich* assoziiert, entscheidet er sich schneller und macht dabei weniger Fehler, wenn diese beiden gedanklich verknüpften Kategorien auf der selben Seite stehen. Er nutzt dabei aus, dass er sich nur für eine der beiden Kategorien die Seite „merken“ muss. Bei jedem Stimulus braucht er nur „entscheiden“, ob er es mit der Kategorie assoziiert, von der er sich die Seite „gemerkt“ hat und diese dann wählen.⁵ Der Unterschied zwischen den Latenzzeiten des kompati-

⁵Letztendlich hilft ihm dieser Mechanismus schon in den Blöcken 1, 2 und 3, da man auch die Zuordnung eines Stimulus' zu einer Kategorie als Assoziation interpretieren kann, auch wenn sie durch die Fehlermeldung aufgezwungen wird.

blen und des inkompatiblen Blocks kann als Maß für die Assoziationen zwischen den kompatiblen Paarungen interpretiert werden.

An dieser Stelle zeigt sich eine vermeintliche Schwäche des IATs. Da er in der Ziel-dimension zwei Kategorien benötigt, gibt es in den kompatiblen Blöcken auch zwei vermeintlich kompatible Paarungen (*Naturwissenschaften/männlich* und *Geisteswissenschaften/weiblich*). Daraus folgt, dass nicht eindeutig klar ist, auf welche Assoziation der Latenzzeitunterschied zurückzuführen ist. Der IAT ermöglicht also nur relative Aussagen über zwei mögliche Assoziationen. Bei einem Testergebnis, welches die Aussage „Im Vergleich zu *Geisteswissenschaften* werden *Naturwissenschaften* eher mit *männlich* assoziiert“ muss auch die Interpretation „Im Vergleich zu *Naturwissenschaften* werden *Geisteswissenschaften* eher mit *weiblich* assoziiert“ in Betracht gezogen werden.⁶

4.1.2 Berechnung des Assoziationsmaßes (*D*-Wert)

Aus den Rohdaten, die man bei einem IAT erhält, lässt sich auf viele verschiedene Arten ein Assoziationsmaß bestimmen. Dabei sind jedoch nicht nur Aspekte der reinen Berechnung zu beachten, sondern bereits die Art der Fehlerrückmeldung im Test und die anschließende Datenaufbereitung sind dafür relevant. Greenwald, Nosek und Banaji (2003) schlagen einen *Improved Scoring Algorithm* vor, der sich in vielen entscheidenden Punkten von dem ursprünglich eingesetzten Algorithmus (vgl. Greenwald et al., 1998) unterscheidet. Bis dahin bestand der Ablauf der meisten Studien, in denen ein IAT eingesetzt wurde, aus folgenden Schritten (vgl. Greenwald et al., 2003, Tabelle 4):

1. Ausschließliche Nutzung der Daten aus den beiden Testblöcken (Block 4 und 7).
2. Unsystematische Löschung von Probanden, die extrem langsam und/oder mit hohen Fehlerraten geantwortet haben.
3. Nichtberücksichtigung der ersten beiden Trials in beiden Blöcken.
4. Ersetzung von Latenzzeiten $< 300\text{ ms}/> 3000\text{ ms}$ durch 300 ms bzw. 3000 ms.
5. log-Transformation aller Werte.⁷
6. Mittelwertbildung in beiden Blöcken.
7. Berechnung der Differenz der beiden Mittelwerte.

Die Autoren vergleichen in einer sehr groß angelegten Studie ($N \approx 1\,200\,000$) verschiedene Statistiken (Mittelwert, Median), verschiedene Transformationen der Statistiken, verschiedene Umgänge mit Übungs- und Fehler-Trials und schließlich verschiedene

⁶Es gibt mehrere Möglichkeiten, diesem „Problem“ zu begegnen, siehe Abschnitt 4.1.3 (**Einfluss der Assoziationskategorien**) und Abschnitt 4.2.

⁷Die Verteilung der Latenzzeiten ist typischerweise extrem linksschief (siehe Abb. 4.1). Mithilfe der log-Transformation korrigiert man dies mit dem Ziel, annähernd normal verteilte Daten zu erhalten.

Ausschlusskriterien für Trials oder sogar Probanden. Im Folgenden wird nur der in der vorliegenden Arbeit eingesetzte Algorithmus vorgestellt, der sich weitestgehend mit den Empfehlungen von Greenwald et al. (2003) deckt und sich auch nur in Details von den in vergleichbaren Studien (Kessels et al., 2006; Nosek et al., 2009; Cvencek et al., 2011) angewendeten Regeln unterscheidet.⁸

Der Umgang mit Fehler-Trials wird schon durch die Zeitmessung während der Durchführung des IATs vorherbestimmt. Bei einer falschen Zuordnung erhält der Proband eine Fehlerrückmeldung. Der nächste Stimulus erscheint erst, wenn er die Eingabe korrigiert hat. Die Zeitmessung für diesen Trial stoppt erst, wenn der Stimulus richtig zugeordnet ist. Dies entspricht einem *Built-In-Error-Penalty* (BIEP). Für jeden Fehler erhält der Proband so eine zusätzliche Strafzeit. Dies ist der Annahme geschuldet, dass in inkompatiblen Blöcken mehr Fehler gemacht werden. Durch eine Strafzeit wird diese höhere Fehlerhäufigkeit in eine höhere mittlere Latenzzeit umgewandelt, damit sich die weitere Auswertung auf die Betrachtung der Zeiten beschränken kann. Die Strafzeit bei dieser Methode ist in der Größenordnung der menschlichen Reaktionszeit, also etwa 300 ms bis 500 ms. Alternativ könnte auch jeder Fehler-Trial durch eine Konstante ersetzt werden (z. B. Blockmittelwert+2SD), was die Korrektur des Fehlers durch den Probanden nicht mehr erfordern würde. Der Einfluss verschiedener Penalty-Strategien auf das Assoziationsmaß ist jedoch gering (vgl. Greenwald et al., 2003, S.213).

Vor jeder weiteren Berechnung werden alle Trials gelöscht, in denen die Latenzzeit größer als 10000 ms ist. Es treten vereinzelt solch extrem hohe Latenzzeiten auf, bei denen davon auszugehen ist, dass der Proband kurzzeitig abgelenkt und anderweitig beschäftigt ist (siehe auch Abb. 4.1). Das Beibehalten solcher Ausreißer würde die Messung zu stark verzerren und in Anbetracht der relativ hohen Trialanzahl pro Proband und Block können diese Einzelfälle bedenkenlos gelöscht werden.

Des Weiteren werden auch alle Fälle (Probanden) gelöscht, bei denen in mehr als 10% der Trials die Latenzzeiten kleiner als 300 ms sind. Bei diesen Probanden kann man davon ausgehen, dass sie den IAT nicht gewissenhaft durchgeführt haben, da sie zu häufig schneller als die durchschnittliche menschliche Reaktionszeit geantwortet haben. Die einzige Möglichkeit den Test abzubrechen besteht für die Probanden darin, sich schnell durch alle Trials „durchzutippen“. Bei der BIEP-Methode ist dies durch schnelles abwechselndes Tippen der Tasten *W* und *P* möglich. Durch das häufige Auftreten der sehr kurzen Latenzzeiten lassen sich solche Fälle identifizieren. Eine Alternative dafür wäre, die Anzahl der Fehler-Trials als Ausschlusskriterium zu nutzen. Allerdings ist es schwierig dafür einen geeigneten Prozentsatz zu finden, da auch bei gewissenhaft durchführenden Probanden viele Fehler denkbar wären. Das Filtern bestimmter Probanden kann auch im Nachhinein vollzogen werden. Das SPSS-Skript auf der Homepage von Millisecond SoftwareTM (Millisecond Software, 2013) berechnet für jeden Proban-

⁸Das SPSS-Skript für diesen Algorithmus kann auf der Homepage von Millisecond SoftwareTM (Millisecond Software, 2013) geladen werden. Dort befindet sich auch eine sehr umfangreiche Testbibliothek.

den die prozentualen Anteile an Trials, die kürzere Latenzzeiten als 300 ms und 400 ms haben.

Für die Berechnung des „neuen“ Assoziationsmaßes werden im Gegensatz zur ursprünglichen Methode nicht nur die Testblöcke (Block 4 und 7), sondern auch die zugehörigen Übungsblöcke berücksichtigt (Block 3 und 6). Greenwald et al. (2003) begründen dies mit der Tatsache, dass die Messungen der Blöcke 3 und 6 nicht nur zufriedenstellend mit den expliziten Messungen korrelieren, sondern teilweise sogar besser als die der Blöcke 4 und 7, was für eine bessere konvergente Validität spricht. Es werden also in allen Blöcken die Stichprobenmittelwerte ($\bar{x}_3, \bar{x}_4, \bar{x}_6, \bar{x}_7$) und die Stichprobenstandardabweichungen (s_3, s_4, s_6, s_7) berechnet. Mit diesen Kennwerten wird ein Assoziationsmaß für die Übungsblöcke (D_a) und eines für die Testblöcke (D_b) nach folgenden Formeln berechnet:

$$\begin{aligned} D_a &= \frac{\bar{x}_6 - \bar{x}_3}{\sqrt{\frac{(N_3-1)s_3^2 + (N_6-1)s_6^2 + \frac{(N_3+N_6)(\bar{x}_3-\bar{x}_6)^2}{4}}{N_3+N_6-1}}}, \\ D_b &= \frac{\bar{x}_7 - \bar{x}_4}{\sqrt{\frac{(N_4-1)s_4^2 + (N_7-1)s_7^2 + \frac{(N_4+N_7)(\bar{x}_4-\bar{x}_7)^2}{4}}{N_4+N_7-1}}}. \end{aligned} \quad (4.1)$$

Aus den beiden Werten wird anschließend das arithmetische Mittel gebildet:

$$D = \frac{D_a + D_b}{2}, \quad (4.2)$$

welches das resultierende Assoziationsmaß (D -Wert) darstellt. Der Umstand, dass ein ungewichteter Mittelwert über die beiden D -Werte gebildet wird, obwohl die Trial-Anzahl in den Blöcken nicht gleich ist ($N_3 = N_6 = 20, N_4 = N_7 = 40$), wird mit der höheren Validität der Übungsblöcke gerechtfertigt (s. o.).

Der Nenner in Formel (4.1) entspricht einer Standardabweichung. Diese Normierung soll Alterseffekte korrigieren, die in Studien mit reaktionszeitbasierten Verfahren auftreten. Bei älteren Probanden sind sowohl die Effekte als auch die Varianzen der Latenzzeiten größer als bei jüngeren (mehr Details siehe Greenwald et al., 2003).

Der D -Wert ähnelt dem bekannten Effektstärkemaß Cohens d mit dem Unterschied, dass bei Cohens d die Standardabweichungen innerhalb der beiden Gruppen/Blöcken zu einer gepoolten Standardabweichung zusammengefasst werden, während beim D -Wert eine gepoolte Standardabweichung über die Trials beider Blöcke hinweg gebildet wird. Beim Vergleich mit d ist D als Assoziationsmaß schließlich überlegen.

Der gesamte Algorithmus für die Datenaufbereitung und Berechnung des D -Wertes ist in Tabelle 4.1 zusammengefasst.

4.1.3 Psychometrische Eigenschaften

Durch die große Bedeutung des IATs für die Psychologie (s. o.) sind in den 15 Jahren seit der ersten Veröffentlichung unzählige Studien veröffentlicht, die sich mit der

4 Messung von impliziten Assoziationen

Tabelle 4.1 Zusammenfassung der Datenaufbereitung und der Berechnung des D-Wertes, wie es Greenwald et al. (2003) empfehlen.

1	Lösche alle Trials mit einer Latenzzeit > 10000 ms.
2	Entferne alle Fälle (Probanden), bei denen in mehr als 10% der Trials die Latenzzeiten < 300 ms sind.
3	Berechne jeweils die gepoolte Standardabweichung für alle Trials der Blöcke 3 & 6 und der Blöcke 4 & 7.
4	Berechne die Mittelwerte der Latenzzeiten in den Blöcken 3, 4, 6 und 7.
5	Berechne die zwei Mittelwertdifferenzen ($\text{Mittelwert}_6 - \text{Mittelwert}_3$) und ($\text{Mittelwert}_7 - \text{Mittelwert}_4$)
6	Teile jede Differenz durch die zugehörige gepoolte Standardabweichung.
7	D = der (ungewichtete) Mittelwert der beiden Quotienten

Anmerkung Übersetzung von Tabelle 3 in Lane et al. (2007).

Eignung des IATs als *Test* im Sinne der klassischen Testtheorie auseinandersetzen. Vor allem hinsichtlich der Validität, insbesondere der internen Validität⁹, gibt es kontroverse Diskussionen darüber, ob der Test das misst, was er für sich beansprucht (automatisch aktivierte, implizite Assoziationen) und ob die kognitiven Vorgänge bei seiner Durchführung hinreichend verstanden sind. Die große Anzahl an in der Praxis angewendeten IATs (siehe z. B. der Überblick von Greenwald, 2012, $N = 59$) und Studien, in denen der Test selbst Untersuchungsgegenstand ist (z. B. Bluemke & Friese, 2006; Greenwald et al., 2003; Mierke, 2004; Nosek, Greenwald & Banaji, 2005; Sriram & Greenwald, 2009), lassen jedoch sehr gut fundierte Aussagen über die Testgüte zu. Das spiegeln auch die Meta-Analysen zur Vorhersage- und Konstruktvalidität wider (Greenwald, Banaji, Poehlmann & Uhlmann, 2009; Hofmann et al., 2005), genauso wie die Review-Artikel zu dem Thema (Cvencek, Greenwald & Meltzoff, 2012; Fazio & Olson, 2003; Gawronski & Conrey, 2004; Lane et al., 2007; Nosek et al., 2007; Schnabel, Asendorpf & Greenwald, 2008).

⁹Unter der Überschrift *Interner Validität* werden in der Literatur zum IAT zwei Dinge diskutiert. 1. Die dem Test zugrundeliegenden psychologischen Prozesse (Gawronski & Conrey, 2004) und 2. Testdesign-Faktoren, welche die Testergebnisse beeinflussen, wie z. B. die Art der Stimuli, Testreihenfolge oder die Kategoriebezeichnungen (Nosek, Greenwald & Banaji, 2007). Der 1. Punkt ist mittlerweile gut untersucht und auf verschiedene Arten modelliert worden. Eine detaillierte Darstellung würde hier zu weit führen. Weitere Informationen und Literaturhinweise dazu gibt es von Lane et al. (2007, S.84) unter *FAQ – What processes underlie IAT effects?*. Der 2. Blickwinkel auf die interne Validität wird unten diskutiert.

Reliabilität

Für die Schätzung der Reliabilität durch die interne Konsistenz der Trials haben sich mehrere Verfahren etabliert (Schnabel, Asendorpf & Greenwald, 2007). Oftmals werden Split-Half-Reliabilitäten über die Differenzzeiten der Blöcke 3/6 und 4/7 oder über deren *D*-Werte (siehe Formel (4.1)) berechnet (z. B. Nosek, 2007). Eine weitere Möglichkeit besteht darin, Differenzzeiten zwischen kompatiblen und inkompatiblen Blöcken für jeden Trial/Stimulus zu berechnen und diese Differenzen als Items zu behandeln. Über diese wird dann die interne Konsistenz Cronbachs α berechnet (z. B. Bluemke & Friese, 2008). Da dieses Maß stark von der Itemanzahl abhängt, können Blöcke über mehrere Differenzen gebildet werden (z. B. 10 Trials). Deren Mittelwerte werden als Einheiten für die α -Berechnung angesehen, um so eine Überschätzung der Reliabilität aufgrund zu vieler Items zu verhindern. Das in dieser Arbeit eingesetzte Verfahren wird detailliert in Abschnitt 10.4 beschrieben.

Die internen Konsistenzen (Split-Half-Korrelationen oder Cronbachs α) liegen für gewöhnlich zwischen .70 und .90 (Nosek et al., 2007). In diesem Punkt ist der IAT anderen impliziten Verfahren weit überlegen, bei denen die Werte oft nahe Null sind und meistens nicht über $\alpha = .50$ hinausgehen (z. B. Bosson, Swann Jr. & Pennebaker, 2000). Die Test-Retest-Reliabilitäten des IATs sind zwar für implizite Messungen zufriedenstellend, jedoch nicht so hoch, wie die internen Konsistenzen. Hofmann et al. (2005) berichten eine mittlere Korrelation von $r = .51$ (Meta-Analyse, $N = 11$) und bei Nosek et al. (2007) ist sie mit $r = .59$ (Median, $N = 9$) in derselben Größenordnung. Gleichzeitig stellen die Autoren dabei nur einen sehr geringen Zusammenhang zur Zeit zwischen Vor- und Nachtest fest. Die große Divergenz zwischen den Reliabilitätsmaßen lassen den Schluss zu, dass der IAT nicht nur zeitlich stabile (*Trait*), „sondern auch situativ aktivierte Assoziationen misst“ (State, Gawronski & Conrey, 2004, S. 123).

Konstruktvalidität

Für die Schätzung der konvergenten Validität werden Korrelationen zwischen IATs und anderen impliziten Verfahren (z. B. Priming-Prozeduren und GNATs, s. o.) herangezogen. Schnabel et al. (2008) bemerken in einem Review-Artikel, dass diese typischerweise sehr schwach sind. Bosson et al. (2000) berichten von Korrelationen im Bereich $-.14 < r < .04$ ns. Auch aktuellere Untersuchungen zur konvergenten Validität von Rudolph, Schröder-Abé, Schütz, Gregg und Sedikides (2008) zeigen nur Korrelationen im Bereich $-.04 < r < .09$ ns. Größere Zusammenhänge lassen sich zwischen einem Selbstkonzept-IAT und der Implicit Association Procedure (IAP Schnabel, Banse & Asendorpf, 2006) mit $r \approx .5$ zeigen, wobei sich beide Tests sehr ähneln.¹⁰ Lane et al. (2007) begründen die anscheinend voneinander unabhängigen Assoziationsmaße

¹⁰Beim IAP wird statt der Tasten ein Joystick benutzt. Bei einer Assoziation des Stimulus' mit sich selbst zieht der Proband den Joystick an sich heran, während er ihn bei Assoziationen mit Anderen wegdrückt.

der impliziten Verfahren vor allem mit deren geringen Reliabilitäten, da sie es schwierig machen, die Nullkorrelationen sinnvoll zu interpretieren.

Einen sehr großen Stellenwert nimmt die Untersuchung des Zusammenhangs zwischen expliziten Messungen und IATs für die Schätzung der Validität ein. Zum einen werden die Zusammenhänge im Sinne einer Konstruktvalidierung interpretiert (Schnabel et al., 2008) und zum anderen werden sie für Aussagen über die Kriteriumsvalidität herangezogen (Lane et al., 2007). Die Tatsache, dass der Zusammenhang nicht immer gleich, bei genauerer Untersuchung jedoch systematisch ist, macht die Validierung in beiden Fällen sehr komplex. In einer Metastudie berichten Hofmann et al. (2005) über Korrelationen von $r = -.25$ bis $.60$. Diese, scheinbar widersprüchliche, weite Spanne veranlassen Fazio und Olson (2003, S. 304) zu der Forderung, die bisherige Fragestellung bei derartigen Untersuchungen zu ändern:

In our view the variability regarding the correspondence between implicit and explicit measures indicates that discussion of whether a relation exists is not very productive. We already know enough to be able to say that the question has no simple answer. That is, the answer is “it depends”. (...) we need to be asking a “when” question: When, under what conditions, and for what kind of people, are implicit and explicit measures related?

Dem Appell folgend wurde in mehreren unabhängigen Studien nach Moderatoren gesucht, welche die Korrelation zwischen IAT und expliziten Selbstbeurteilungen beeinflussen. Im Folgenden werden exemplarisch die Ergebnisse von zwei größeren Projekten berücksichtigt. In der bereits erwähnten Meta-Analyse (Hofmann et al., 2005) werden 126 unabhängige Studien ($N > 9000$) auf 10 mögliche Moderatorvariablen hin untersucht, mit folgendem Ergebnis: Die Korrelation zwischen den beiden Messungen ist höher wenn (a) der Grad der Spontanität zu Selbstbeurteilung höher und (b) die konzeptuelle Ähnlichkeit¹¹ beider Messungen größer ist. Außerdem ist der Zusammenhang (c) von der Art der Stimuli abhängig. Pronomen korrespondieren im Gegensatz zu allen anderen Arten von Stimuli (z. B. Bilder, Namen usw.) bei der Zieldimensionen eher mit geringen Korrelationen, während in der Attributdimension bewertende Substantive höhere Korrelationen verursachen als Adjektive. Obwohl ein Hauptziel von impliziten Verfahren ist, die Messung unabhängig von sozialer Erwünschtheit durchzuführen (siehe oben), konnte keine Evidenz dafür festgestellt werden, dass das Themengebiet des IATs (z. B. Stereotypen, Einstellungen gegenüber Gruppen oder gegenüber Konsumprodukten), und damit der Grad der sozialen Erwünschtheit, einen Einfluss auf die Korrelation hat.

In einer weiteren (Primär-)Studie ($N \approx 6800$) kann Nosek (2005) (vgl. auch Nosek, 2007) weitere Moderatoren identifizieren. Die Implizit-Explizit-Korrelation ist höher,

¹¹Der IAT spricht eher affektive Prozesse an und misst Assoziationen relativ bzgl. bipolaren Kategorien. Die Korrelation kann erhöht werden, wenn im expliziten Test ebenfalls eher affektive als kognitive Prozesse aktiviert werden und (statt absoluter) relative Bewertungen vorgenommen werden.

wenn (d) der Hang zur Selbstdarstellung der Probanden geringer ist, (e) die wahrgenommene Individualität¹² größer ist und (f) je vertrauter der Proband mit der Bewertung der Kategorien ist. So steigt der Zusammenhang zwischen den beiden Messungen wenn die Zielkategorien (Personen, ethnische Gruppen, Parteien usw.) für den Probanden eine hohe Relevanz haben oder wenn der Proband sich (im oben gewählten Beispiel) schon oft Gedanken über den Zusammenhang von *Geschlecht* und *Naturwissenschaften* gemacht hat. Außerdem hängt die Korrelation (g) von der individuellen Stärke der Assoziation und (h) von dem Grad der Bipolarität der Kategorien ab. Je größer die Assoziationen und je gegensätzlicher die beiden Kategorien sind, desto stärker die Korrelation. So kann eine nicht sinnvoll gewählte Vergleichskategorie den Zusammenhang negativ beeinflussen.

Obwohl die berichteten mittleren Korrelationen mit $Mdn\ r = .22$ (Lane et al., 2007), $r = .19$ (Hofmann et al., 2005) und $Mdn\ r = .48$ (Nosek, 2007) teilweise nur gering sind, kann von einem klaren positiven Zusammenhang und damit konvergenter Validität ausgegangen werden (vgl. Gawronski & Conrey, 2004, Fazit). Das trifft auch für die diskriminante Validität zu, da der IAT nicht mit expliziten und impliziten Messungen anderer Dimensionen korreliert (vgl. Gawronski & Conrey, 2004; Lane et al., 2007). Diese Aussagen werden auch durch die Ergebnisse einer Multitrait-Multimethod-Analyse von Nosek und Smyth (2007) gestützt, die simultan Evidenz sowohl für konvergente als auch diskriminante Validität liefert. Gleichzeitig zeigen die Strukturgleichungsmodelle der Analyse einen besseren Fit, wenn implizite und explizite Messungen als zwei separate Konstrukte angenommen werden. Daraus lässt sich schlussfolgern, dass diese Konstrukte unterschiedlich sind, auch wenn die Assoziationsmaße der IATs und der expliziten Methoden miteinander korrelieren (vgl. auch Lane et al., 2007).¹³

Kriteriumsvalidität

In einer Meta-Analyse ($N > 13000$, 122 Berichte/Veröffentlichungen, 184 unabhängige Stichproben) zur prognostischen Validität kommen Greenwald et al. (2009) zu dem Schluss, dass der IAT in bestimmten Zieldimensionen in der Lage ist, bessere Vorhersagen über das Verhalten der Probanden zu treffen, als explizite Selbstausskünfte. In Bereichen mit hoher sozialer Brisanz (z. B. *Rassismus*)¹⁴ ist die Korrelation der IAT Messungen mit den Kriterien ($\bar{r}_{ICC} = .24$)¹⁵ signifikant höher als die der expliziten Maße ($\bar{r}_{ECC} = .12$)¹⁶, inkrementelle Validität). Zugleich ist genau in diesen Bereichen die prognostische Validität der expliziten Messungen gering. Diese haben dagegen in Be-

¹²d. h. „(...) die wahrgenommene Diskrepanz zwischen der eigenen Einstellung und der anderer Personen“ (Gawronski & Conrey, 2004, S.120)

¹³„Implicit and Explicit Attitudes Are Related but Distinct Constructs“ (Untertitel von Nosek & Smyth, 2007)

¹⁴alle Bereiche, die von hoher sozialer Erwünschtheit beeinflusst werden

¹⁵ICC: IAT-criterion correlation

¹⁶ECC: Explicit-criterion correlation

reichen wie *politische Vorlieben* oder *Konsumverhalten* höhere Vorhersagekraft als die IATs, obwohl die Korrelationen mit den Kriterien beider Messungen höher als bei den sozial brisanten Themen ist. Mit anderen Worten: Der IAT hat in bestimmten Bereichen eine höhere Vorhersagevalidität als explizite Messungen. Über alle Domänen hinweg ist sie zufriedenstellend hoch, auch wenn sie (wie die der expliziten Messungen) stark variiert. Somit verschaffen beide Messungen zusammen ein Zugewinn an prognostischer Validität gegenüber dem Gebrauch von einer allein.

Weitere Studien, die Belege für nomologische und Übereinstimmungsvalidität (Known-Groups Validity) liefern, werden von Lane et al. (2007) aufgeführt. Zusammenfassend bemerken die Autoren des Review-Artikels: „In short, the IAT can predict group membership based on theoretically predicted patterns of ingroup attitudes and identification, correlates with (but is distinct from) explicit measures of associated constructs, and successfully predicts judgments and behaviors.“ (S. 75).

Einfluss der Assoziationskategorien

Genauso wie Fragebögen oder andere Formen der expliziten Selbstbeurteilung auch, beeinflusst das Design des IATs das Antwortverhalten. Analog zur Skalenbeschriftung bei einer Likert-Skala sollten auch die Zielkategorien und deren Bezeichnungen bewusst und unmissverständlich gewählt werden. Auf jeden Fall sollten Konfundierungen vermieden werden, die durch zweideutige Bezeichnungen und Stimuli entstehen könnten (siehe Fußnote 3 auf Seite 61 und **Einfluss der Stimuli**). Bei der Wahl der Assoziationskategorien ist außerdem zu beachten, dass sie dem Test einen Kontext geben, der ebenfalls Einfluss auf den *D*-Wert hat. Mitchell, Nosek und Banaji (2003) vergleichen die Ergebnisse zweier IATs mit identischen Attributkategorien (*good/bad*) und identischen Stimuli. Diese sind zum einen beliebte schwarze¹⁷ Sportler und zum anderen unbeliebte weiße Politiker. Im ersten IAT lautet die Zieldimension *Occupation*¹⁸ (*athlete/politician*), was im Mittel eine Assoziation von *athlete* mit *good* bzw. *politician* mit *bad* hervorruft. Der zweite IAT hat die Zieldimension *Race*¹⁹ (*black/white*), mit dem Ergebnis, dass nun eher *black* mit *bad* gedanklich verbunden wird. In beiden Fällen waren die Stimuli die selben schwarzen Sportler, welche aber je nach Kontext (Zieldimension *Occupation* oder *Race*) mit unterschiedlichen Attributen assoziiert werden.

Einfluss der Stimuli

Wenn die Kategorien und deren Bezeichnungen mit den Skalenbeschriftungen einer Likert-Skala vergleichbar sind, dann entsprechen die Stimuli des IATs den Items eines Fragebogens. Dementsprechend hoch ist auch ihr Einfluss auf das Messergebnis.

¹⁷Als Bezeichnung für *afro-amerikanisch* wird hier die Übersetzung des in der englischsprachigen Literatur und in den IATs verwendeten Begriffs *Black* verwendet.

¹⁸engl.: Beruf

¹⁹In Nordamerika ist der Begriff in der Wissenschaft und bei Behörden (noch) üblich.

Neben der möglichen Verwechslung, ob ein Stimuli nun zu der einen oder der anderen Ziel-/Attributkategorie gehört (siehe Fußnote 3 auf Seite 61), ist auch eine Konfundierung über Ziel- und Attributkategorie hinweg möglich (*cross-category association* Bluemke & Frieze, 2006). In einer Variation des Experiments zur Untersuchung der Kontexteffekte der Zielkategorien (siehe oben) verwenden Mitchell et al. (2003) im IAT mit der Zieldimension *Race* zunächst beliebte schwarze und unbeliebte weiße Personen und danach unbeliebte schwarze und beliebte weiße Personen als Ziel-Stimuli, was einen großen signifikanten Unterschied in der Bewertung der Dimension *Race* in beiden Tests nach sich zieht. Dies bestätigen auch die elaborierteren Studien von Bluemke und Frieze (2006) zum Einfluss der Stimuli. In einem Ost/West-IAT werden die Assoziationen von westdeutschen Probanden bezüglich der Attribute *positiv* und *negativ* gemessen. Dabei werden systematisch konfundierte Ziel- und Attributstimuli verglichen: positiv konnotierte Ost- und West-Stimuli (z. B. *Ostsee* und *Freiheit*), negativ konnotierte Ost- und West-Stimuli (z. B. *Stasi* und *RAF*), mit *West* konnotierte Positiv- und Negativ-Stimuli (z. B. *dynamisch* und *überheblich*) und schließlich mit *Ost* konnotierte Positiv- und Negativ-Stimuli (z. B. *natürlich* und *ausländerfeindlich*). Außerdem werden für jede Kategorie neutrale Stimuli im Vergleich berücksichtigt. Die Ergebnisse zeigen eine starke Abhängigkeit des *D*-Wertes davon, ob die Attribut-Stimuli, die Ziel-Stimuli oder beide über Kreuz assoziiert werden. Die Werte reichen von $D = -0.81$ (starke Assoziation pro-Ost) bis $D = 2.70$ (sehr starke Assoziation pro-West). Im oben gewählten Beispiel (Naturwissenschaften/Geschlecht-IAT, siehe Abschnitt 4.1.1) tritt eine solche *cross-category association* durch die Verwendung von Personen als Ziel-Stimuli auf, da diese immer zugleich auch einem Geschlecht, also der Attributdimension zugeordnet werden können. Daher sollten Namen oder Bilder von Personen bei Geschlechts-IATs nur als Stimuli für die Dimension *Geschlecht* verwendet werden.

Lane et al. (2007) machen noch auf weitere mögliche Konfundierungen aufmerksam. Es könnten z. B. alle Naturwissenschafts-Stimuli Fremdworte sein oder alle weiblich-Stimuli weniger als vier Buchstaben haben, wodurch die Probanden eventuell nach einer für den Test irrelevanten Regel sortieren würden. Ferner raten sie von Negationen ab, wie *unintelligent*, weil sie schwerer korrekt zuzuordnen sind.

Im Unterschied zu Items eines Fragebogens hat die Anzahl der Stimuli in den verschiedenen Blöcken keinen großen Einfluss auf die Testgüte. Bisher gibt es keine Hinweise dafür, dass z. B. die Reliabilität durch mehr Trials gesteigert werden kann (Lane et al., 2007). Lediglich die Anzahl an Stimuli sollte zwei pro Kategorie nicht unterschreiten (Nosek et al., 2005).

Weitere beeinflussende Faktoren

Die folgenden Aspekte haben alle mehr oder weniger Einfluss auf den *D*-Wert oder die Testgüte des IATs. Sie werden jedoch nur aufgezählt und nicht detailliert erläutert. So soll deutlich werden, wie umfangreich das Spektrum der Faktoren ist, die sich auf die

Ergebnisse der in dieser Arbeit eingesetzten IATs auswirken können. Für ausführlichere Informationen wird auf die jeweilige Literatur verwiesen.

Bezüglich der Reihenfolge der kombinierten Blöcke können Verschiebungen der *D*-Werte in Richtung der ersten Paarung beobachtet werden. Dem kann durch eine Erhöhung der Trialanzahl in Block 5 oder bei Gruppenvergleichen durch die Randomisierung der Reihenfolge entgegengewirkt werden (Nosek et al., 2005). Bei mehreren IAT-Messungen nacheinander können Übungseffekte auftreten, so dass mit zunehmender Anzahl an durchgeführten Tests die IAT-Effekte geringer werden. Andererseits werden mit zunehmendem Alter der Probanden die IAT-Effekte größer. Der in Abschnitt 4.1.2 beschriebene Algorithmus für den *D*-Wert reduziert beide Effekte, jedoch sollte der für die Untersuchung bedeutsamste IAT zu Beginn der Messung durchgeführt werden (Greenwald et al., 2003; Nosek et al., 2007). Die Reihenfolge von impliziter und expliziter Messung hat wiederum keinen systematischen Einfluss auf die Ergebnisse (Hofmann et al., 2005; Nosek et al., 2005; Nosek, 2005). Wichtig für vergleichende Studien, vor allem im schulischen Kontext, ist der Befund, dass der Einfluss der kognitiven Fähigkeit auf das IAT-Ergebnis eliminiert werden kann. Wright und Meade (2012) können belegen, dass die nach der ursprünglichen Methode (Greenwald et al., 1998) berechneten IAT-Effekte signifikant mit der kognitiven Fähigkeit korrelieren, während dies für die *D*-Werte (Greenwald et al., 2003) nicht gilt.

4.1.4 Bisheriger Einsatz im Kontext *Naturwissenschaften* und *Schule*

Neben den vielfältigen Einsätzen des IATs in unterschiedlichen Bereichen der Psychologie (siehe Abschnitt 4.1) gibt es auch Studien, die ihn bei Fragestellungen und Kontexten der Bildungs- und Erziehungswissenschaften bzw. der pädagogischen Psychologie verwenden. Stellvertretend dafür werden hier vier solcher Studien aufgeführt, deren Ergebnisse eine Bedeutung für die vorliegende Arbeit haben.

In einer frühen Studie untersuchen Nosek et al. (2002) mithilfe von IATs den Zusammenhang zwischen der Einstellung zur Mathematik²⁰, der Identifikation mit Mathematik²¹, der Mathematik/Geschlechts-Stereotypisierung²² und der Geschlechtsidentität²³ bei $N \approx 80$ Studierenden der Psychologie. Die inhaltlichen Ergebnisse der Studie sind bereits in Abschnitt 3.3 aufgeführt. Bezüglich des eingesetzten Testverfahrens kann jedoch zusammengefasst werden, dass der IAT bereits erfolgreich zur Messung von Assoziationen eines Wissenschaftsgebiets (Mathematik) eingesetzt wurde, dessen Image zumindest im deutschsprachigen Raum ähnlich wie das von *Physik* ist. Die Operationalisierungen der oben aufgezählten Konstrukte durch die jeweiligen impliziten Asso-

²⁰Assoziation von *math/arts* und *pleasant/unpleasant*

²¹...*math/arts* und *self/other*

²²...*math/arts* und *male/female*

²³...*male/female* und *self/other*

ziationen (siehe Fußnoten) ist nach Ansicht der Autoren nicht nur zulässig, sondern sie ermöglicht auch eine höhere Vorhersagekraft²⁴ als die expliziten Messungen.

In einer weiteren Studie gehen Nosek et al. (2009) der Frage nach, inwieweit nationale Unterschiede bei impliziten Gender/Science-Stereotypisierungen die Unterschiede zwischen Männern und Frauen in Mathematik- und Science-Leistungstests²⁵ vorhersagen können (siehe auch Abschnitt 3.3). Dieser Studie liegen Daten aus $N \approx 300000$ absolvierten IATs mit den Zieldimensionen *science/liberal arts* und den Attributdimensionen *male/female* zugrunde. Auch hier wird durch die implizit gemessenen Stereotypen mehr Varianz bei den Unterschieden zwischen Männern und Frauen in den Leistungstests aufgeklärt (19% und 24%) als durch die expliziten (2% und 1%), was eine höhere Vorhersagevalidität bedeutet.

Cvencek et al. (2011) messen die Mathematik/Geschlechts-Stereotypisierung, Geschlechtsidentität und die Identifikation mit Mathematik bereits bei Grundschulern. Neben dem inhaltlichen Ergebnis, dass die von Erwachsenen bekannten Assoziationen schon in diesem Alter auftreten, wird außerdem indirekt gezeigt, dass der IAT auch erfolgreich in jüngeren Zielpopulationen (Alter 6 bis 11 Jahre) ohne Einbußen bei der Testgüte eingesetzt werden kann. So waren z.B. die internen Konsistenzen der IATs zwischen $\alpha = .74$ und $\alpha = .93$, während die expliziten Messungen der Mathematik/Geschlechts-Stereotypisierung bei $\alpha = .03$ lagen. Weitere Einsätze von IATs bei Kindern findet man bei Dunham, Baron und Banaji (2006) und Rutland, Cameron, Milne und McGeorge (2005).

Kessels et al. (2006) setzen fünf IATs (Dimensionen *identification*, *attitude*, *heteronormy*, *masculinity* und *difficulty*) für die Messung des Images von Physik ein (siehe Abschnitt 3.2). Die Autorinnen stellen fest, dass sich die impliziten Assoziationen in einigen Dimensionen nach bereits kurzer Intervention ändern. Das ist ein Beleg dafür, dass sich der IAT in den für die vorliegende Arbeit relevanten Dimensionen dazu eignet, Treatmenteffekte aufzudecken. Die Effektstärke für den Unterschied der *D*-Werte vor und nach der Intervention lag im Falle eines IATs bei $d = 1.0$ (großer Effekt, siehe Tabelle C.1 im Anhang).

4.2 Single Target IAT

Der IAT benötigt in der Ziel- und der Attributdimension immer eine Vergleichskategorie. In dem Beispiel oben (Abschnitt 4.1.1) kann nicht eindeutig von einer Assoziation von *Naturwissenschaften* mit *männlich* ausgegangen werden, genauso wenig, wie eindeutig von einer Assoziation von *Geisteswissenschaften* mit *weiblich* ausgegangen werden kann. Vielmehr ermöglicht der IAT Aussagen wie: “Im Vergleich zu Geisteswissenschaften werden Naturwissenschaften eher mit männlich, als mit weiblich assoziiert.“ Es handelt sich also immer um eine relative Messung.

²⁴ „predictive power“ (Nosek et al., 2002, S. 57)

²⁵ Grundlage sind die Daten von TIMSS 2003 (Martin, Mullis & Chrostowsk, 2004).

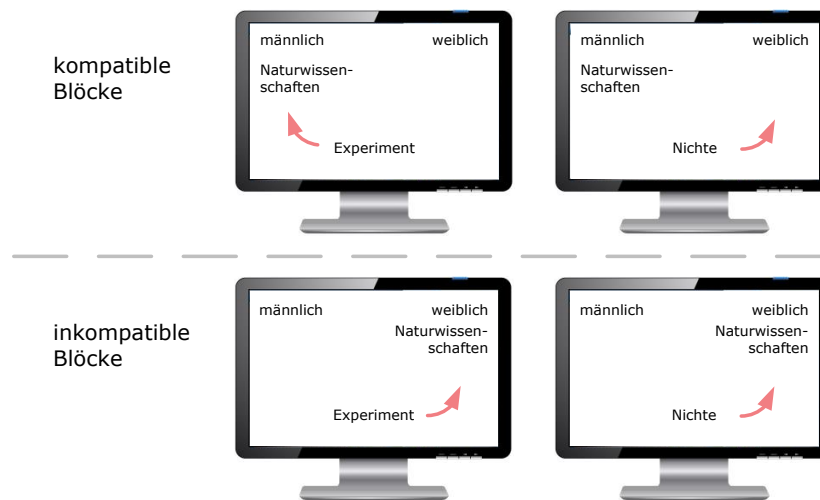


Abb. 4.3 Zwei beispielhafte Trials eines ST-IATs in den kompatiblen Blöcken 2 und 3 (oben) und inkompatiblen Blöcken 4 und 5 (unten).

Wigboldus, Holland und van Knippenberg (2006) stellen eine alternative Prozedur vor, die Assoziationen ohne die sonst notwendige Vergleichskategorie messen soll, den Single Target IAT (ST-IAT). Wie der Name deutlich machen soll, handelt es sich beim ST-IAT um ein dem IAT sehr ähnliches Verfahren, jedoch mit dem Unterschied, dass nur eine Zielkategorie (single target) verwendet wird. Somit besteht die Möglichkeit, absolute Assoziationen zu messen, z. B. in Dimensionen, bei denen es keine sinnvollen Vergleichskategorien gibt. Die bis zu dem Zeitpunkt vorgenommenen Versuche, dies durch getrennte *D*-Wert-Berechnungen für die Stimuli der beiden Zielkategorien (vgl. Nosek et al., 2005, Studie 1) zu erreichen, brachten keine zufriedenstellenden Ergebnisse. Auch die Verwendung neutraler Vergleichskategorien ist nicht ratsam, da es erstens schwierig ist, eine wirklich neutrale Kategorie zu wählen und zweitens die kognitiven Prozesse bei der Durchführung unverstanden sind (vgl. Lane et al., 2007).

4.2.1 Ablauf eines ST-IATs und Berechnung des *D*-Werts

Der ST-IAT unterteilt sich nur in fünf Blöcke. Die beiden Übungsblöcke für die nun allein stehende Zielkategorie fallen weg. Der erste Block ist wie beim IAT ein Übungsblock mit 20 Trials, nun jedoch für die beiden Attributkategorien. Er hat dieselbe Funktion, wie der entsprechende Block beim IAT: kennenlernen der Stimuli und vertraut werden mit der Sortieraufgabe. Der zweite Block ist bereits ein kombinierter Block, je nach Randomisierung kompatibel oder inkompatibel. Unter den beiden Attributkategorien steht nun die Zielkategorie auf einer der beiden Seiten am oberen Bildschirmrand (siehe Abbildung 4.3). Die ersten 20 Trials dienen wieder der Übung, dann beginnt der dritte Block und damit der erste Testblock (35 Trials). Anschließend wechselt die Zielkategorie die

Seite. Die ersten 20 Trials wird die neue Konstellation geübt (vierter Block) und die folgenden 35 Trials stellen den fünften Block (zweiten Testblock) dar.

Die Aufgabe der Probanden ist identisch mit der des IATs. Mithilfe von zwei Tasten sollen die Stimuli so schnell und mit so wenigen Fehlern wie möglich auf die richtige Seite sortiert werden. Auch die dem Test zugrunde liegende Annahme ist gleich:

Die Sortieraufgabe fällt den Probanden in den kompatiblen Blöcken leichter, so dass sie diese schneller und mit weniger Fehlern durchführen. Aus dem Reaktionszeitunterschied wird auf eine Assoziation zwischen den entsprechenden Kategorien geschlossen.

Ein weiterer, eher technischer Unterschied zum IAT besteht in dem Verhältnis des Auftretens der Stimuli der drei Kategorien. Die Anzahl der möglichen Stimuli ist für alle Kategorien gleich und liegt z. B. bei fünf Stimuli. Damit jedoch das Verhältnis zwischen rechtem und linkem Tastendruck ausgeglichen ist, wählen Wigboldus et al. (2006) ein Verhältnis von 5 : 5 : 10, d. h. die beiden auf einer Seite stehenden Kategorien sind in den Trials nur halb so oft wie die alleinstehende Kategorie vertreten und somit werden genauso viele Wörter nach links wie nach rechts sortiert. Dies verursacht jedoch eine Konfundierung, da im kompatiblen und inkompatiblen Block je eine der beiden Attributkategorien unterrepräsentiert ist. Würde man ein ausgeglichenes Verhältnis aller Kategorien wählen, würden die kompatible und inkompatible Paarung jedoch mit der Anzahl der Rechts- bzw. Links-Antworten konfundieren. Ein Zeitunterschied könnte dann auch auf verschiedene Reaktionszeiten der linken und rechten Hand zurückzuführen sein. Einen Kompromiss schlagen Bluemke und Frieze (2008) mit dem Verhältnis 2 : 2 : 3 und Karpinski und Steinman (2006)²⁶ mit dem Verhältnis 7 : 7 : 10 vor.

Für die Berechnung eines Assoziationsmaßes stehen beim ST-IAT die gleiche Anzahl und Art an Daten wie beim IAT zur Verfügung. Damit bleiben auch sämtliche Möglichkeiten zum Umgang mit ihnen erhalten, wie z. B. die Löschung von Ausreißern und nicht gewissenhaft teilnehmenden Probanden, die Verrechnung von Fehlertrials sowie Möglichkeiten der Transformationen von Latenzzeiten. Daher kann analog zu dem Verfahren, wie es in Abschnitt 4.1.2 vorgestellt ist, auch für den ST-IAT ein *D*-Wert berechnet werden. Dieser kann nun aber als absolute Assoziation²⁷ interpretiert werden und ist nur bedingt mit dem *D*-Wert eines IATs vergleichbar. Karpinski und Steinman (2006) verwenden erfolgreich die Differenz der *D*-Werte zweier ST-IATs für den Vergleich mit dem Wert des entsprechenden IATs. In dem in Abschnitt 4.1 gewählten Beispiel wäre also die Differenz zwischen dem Naturwissenschaften-STIAT-Wert und dem

²⁶Die Autoren verwenden einen Single Category IAT (SC-IAT), der sich in einigen kleinen Details, hauptsächlich der Umgang mit Fehlertrials, vom hier vorgestellten ST-IAT unterscheidet. Zur besseren Lesbarkeit wird im Folgenden in beiden Fällen die Bezeichnung Single Target IAT (ST-IAT) verwendet.

²⁷Dies ist zumindest insofern sinnvoll, als dass der Unterschied zum vergleichenden Charakter des IATs verdeutlicht wird. Karpinski und Steinman (2006) machen darauf aufmerksam, dass es absolute Einstellungen oder Assoziationen nicht gibt, da sie immer eine vergleichende Bewertung irgend einer Art benötigen.

Geisteswissenschaften-STIAT-Wert mit dem IAT-Wert vergleichbar, der beide Zielkategorien gegenüberstellt.

4.2.2 Psychometrische Eigenschaften

Obwohl der ST-IAT seltener eingesetzt wird als der IAT und dadurch bei Weitem nicht derart umfassende Studien zu dessen Testgüte vorliegen, können fundierte Aussagen über dessen psychometrische Eigenschaften gemacht werden.

Reliabilität

Für die Schätzung der Reliabilität durch die interne Konsistenz werden beim ST-IAT im Wesentlichen dieselben Verfahren verwendet wie beim IAT (siehe Abschnitt 4.1.3, Reliabilität, S. 67). Sie liegt im direkten Vergleich für gewöhnlich etwas unterhalb der des entsprechenden IATs. Karpinski und Steinman (2006) berichten von Werten zwischen $.55 \leq r \leq .85$ ²⁸ und Bluemke und Frieze (2008) ermitteln Werte zwischen $.66 \leq r \leq .88$. Damit sind die internen Konsistenzen noch immer zufriedenstellend, vor allem im Vergleich mit anderen impliziten Verfahren (Bosson et al., 2000). Die Test-Retest-Reliabilität²⁹ liegt in einer beispielhaften Studie im Bereich $.32 \leq r \leq .62$ ($Mn\ r = .48$ Bluemke & Frieze, 2008), was zwar nicht zufriedenstellend, jedoch in der gleichen Größenordnung wie die des IATs ist.

Validität

Belege für die konvergente Validität von ST-IATs liefert die mittlere Korrelation von fünf Tests mit den entsprechenden expliziten Äußerungen (Bluemke & Frieze, 2008). Die impliziten und expliziten Einstellungen zu den deutschen Bundestagsparteien (CDU/C-SU, FDP, SPD, Die Grünen und PDS) korrelierten durchschnittlich (zwei Studien) mit $r = .43$, was einem mittleren Effekt entspricht (siehe Tabelle C.1 im Anhang). Damit liegt sie höher als die mittlere implizit-explizit-Korrelation bei vergleichbaren IATs, auch wenn die Autoren darauf aufmerksam machen, dass dieser Wert aufgrund der bekannten Moderatoren und Kontexteffekte (siehe Abschnitt 4.1.3) eventuell eine obere Abschätzung der konvergenten Validität ist. Zusätzliche Belege liefern die Korrelation (mittlere Effekte) mit einem IAT (*links/rechts*³⁰ und *positiv/negativ*). Dieselbe Studie liefert auch Indizien für die diskriminante Validität der ST-IATs. Die einzelnen Testwerte korrelieren nicht miteinander, wenn die beiden Parteien unterschiedlichen Flügeln (*links/rechts*) des Bundestages angehören. Die automatische Assoziation/Bewertung der Partei findet also eher auf einer höheren, ideologischen Ebene statt und die Unterschiede

²⁸mittlere Interkorrelation der *D*-Werte von drei gleich großen Testteilen

²⁹nach Minderungskorrektur

³⁰bezogen auf das politische Lager bzw. die Sitzplatzverteilung im Bundestag

der Bewertung zwischen den Parteien können nicht alle aufgelöst werden. Je nach Kontext oder Fragestellung müsste die Einschätzung erfolgen, ob eine solche diskriminante Validität ausreichend ist.

Zu der gleichen Aussage bzgl. der Konstruktvalidität kommen auch Karpinski und Steinman (2006). Mit zwei ST-IATs messen sie die Einstellung zu *Coke*³¹ und *Pepsi*. Aus der Differenz beider *D*-Werte bilden sie den eines *Soda*-ST-IATs. Der *Coke*-ST-IAT korreliert mit den expliziten Einstellungen zu der Marke, nicht jedoch mit denen zu der Marke *Pepsi*. Dementsprechend umgekehrt verhält es sich mit dem *Pepsi*-ST-IAT. Auch der *Soda*-ST-IAT korreliert nur mit den expliziten Einstellungen zu Sodageetränken. Die einzelnen ST-IATs messen also spezifische Einstellungen zu den Marken, aber keine allgemeine Einstellung zu kolahaltigen Getränken. Zusätzlich belegen die Autoren den Tests auch prognostische Validität. Regressionsanalysen mit den *D*-Werten als Prädiktor und die Wahl einer Marke als abhängige Variable zeigen, dass sowohl der *Coke*- als auch der *Pepsi*-ST-IAT die Wahl einer Marke vorhersagt. In einer anderen Teilstudie vergleichen Karpinski und Steinman (2006) einen Selbstwert-ST-IAT mit einem Selbstwert-IAT. Der ST-IAT korreliert mittel bis stark mit dem expliziten Selbstwert, jedoch nur schwach mit dem IAT. Das lässt darauf schließen, dass ein anderes Selbstwertkonstrukt bei nur einer Zielkategorie (*self* und *good/bad*) angesprochen wird als beim vergleichenden IAT (*self/other* und *good/bad*).

Indizien für die prognostische Validität von ST-IATs können auch in Studien gefunden werden, in denen der Test eingesetzt wird, ohne das explizite Ziel, den Test auf seine Testgüte zu untersuchen. Dotsch und Wigboldus (2008) zeigen, dass das Verhalten³² durch implizite Vorurteile gegenüber bestimmter ethnischer Gruppen vorhergesagt werden kann. Sie setzen dafür einen ST-IAT mit marokkanischen Namen als Zieldimension und *positiv/negativ* als Attributkategorien ein. Auch bei Holland, Vries, Hermesen und Knippenberg (2012) sagt die mit einem ST-IAT gemessene Einstellung gegenüber Blutspenden unter bestimmten Bedingungen die Bereitschaft für zukünftige Blutspenden voraus.

Die zahlreichen Faktoren, die das IAT-Ergebnis beeinflussen können, müssen natürlich auch beim ST-IAT beachtet werden. Konfundierungen, die durch ungünstige Wahl der Assoziationskategorien und Stimuli auftreten, sollten ausgeschlossen werden. Den Reihenfolge-Effekten kann genauso begegnet werden, wie es in Abschnitt 4.1.3 beschrieben ist. Der Alterseffekt, der Übungseffekt bei mehreren ST-IATs hintereinander und der Einfluss der kognitiven Fähigkeit auf die IAT-Messung wird durch die Verwendung des *D*-Wert-Algorithmus gemindert und teilweise eliminiert.

³¹Coca Cola

³²gehaltener Abstand zu einer virtuellen Figur

4.3 Bedeutung für die vorliegende Arbeit

Der IAT ermöglicht die Messung impliziter Assoziationen jeglicher Art. Er ist ein etabliertes Testinstrument, welches in zahlreichen Studien mit sehr unterschiedlichen Kontexten eingesetzt wird. Der Grund dafür liegt gerade in dem großen Anwendungsbereich, aber auch in seinen Vorteilen gegenüber anderen impliziten Verfahren, die in den vergleichsweise hohen Effektstärken und guten psychometrischen Eigenschaften liegen. Die Validität des IATs wird kontrovers diskutiert, was jedoch dazu führte, dass sie sehr ausführlich untersucht wurde und der Test diesbezüglich gut erforscht ist. In vielen Fällen ist sie besser als bei expliziten Testverfahren und man kann davon ausgehen, dass ein gemeinsamer Einsatz eines IATs mit expliziten Methoden immer eine höhere Varianzaufklärung zur Folge hat. Der IAT wurde erfolgreich zur Messung von naturwissenschaftsbezogenen Assoziationen und im Kontext Schule eingesetzt. Dies gilt vor allem auch für jene Dimensionen, die Aussagen über das Image eines Schulfaches erlauben. In diesen Fällen konnten auch Veränderungen als Folge einer Intervention mit dem IAT gemessen werden. Außerdem lässt sich der Test auch schon bei Grundschulern einsetzen, bei denen mit seiner Hilfe die typischen Stereotypisierungen bereits identifiziert werden konnten.

Eine Weiterentwicklung zum IAT ist der Single Target IAT, der den Vorteil aufweist, keine vergleichende Zielkategorie zu benötigen. Im Kontext der vorliegenden Arbeit bedeutet dies, dass kein zweites Schulfach zur Gegenüberstellung mit Physik festgelegt werden muss. Seine psychometrischen Eigenschaften sind größtenteils zufriedenstellend, jedoch schlechter als die des IATs.

Der IAT und der ST-IAT stellen die im Rahmen der vorliegenden Arbeit verwendeten Instrumente dar, die zur Messung von impliziten Assoziationen zur Physik dienen. Begleitet werden sie von expliziten Verfahren, die zur zusätzlichen Validierung dienen. Die Tests dienen damit der Erhebung des größten Teils der in Abschnitt 3.4 genannten abhängigen Variablen, dessen Beeinflussung durch die phänomenbasierte Anfangsoptik untersucht werden soll.

Balanced Identity Theory

Die *Balanced Identity Theory* (BIT) ist die theoretische Grundlage für den Schluss von (impliziten) Assoziationen auf Konstrukte wie *Einstellungen*, *Stereotypen*, *Selbstwert* und *Selbstkonzept*. Obwohl das ursprüngliche Ziel nicht die Vereinheitlichung bereits bestehender Theorien war, wird sie trotzdem zunächst als eine „unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept“ (Greenwald et al., 2002) vorgestellt, die viele Elemente der etablierten Kongruitäts-, Dissonanz- und Balance-Theorien aufgreift¹. Die BIT zählt damit zu der Gruppe der Konsistenztheorien. Angelehnt an das Design zu ihrer empirischen Überprüfung, dem *Balanced Identity Design* (BID), geben Cvencek et al. (2012) ihr den kürzeren und prägnanteren Namen.

Ursprünglich lag der Zweck der BIT in dem Versuch, die IAT-Ergebnisse der ersten Jahre zu beschreiben und zu verstehen. Greenwald et al. (2002) nennen als ein beispielhaftes Experiment die implizite Messung, in der Frauen *female* in derselben Ausprägung mit *strength* assoziiert haben, wie *male* mit *strength*. Eigentlich wurde bei den Frauen dieselbe geschlechtsstereotype Assoziation ($\text{male} \hat{=} \text{strength}$) wie bei den Männern erwartet. Eine mögliche Erklärung liegt im hohen Selbstwert der Frauen. Demnach ist die *female-strength* Assoziation eine Konsequenz von bestehenden *female-self*, *self-positiv* und *positiv-strong* Assoziationen. Die BIT beschreibt derartige Verkettungen und ist somit eine Post-Hoc-Theorie zur Erklärung von IAT-Ergebnissen.

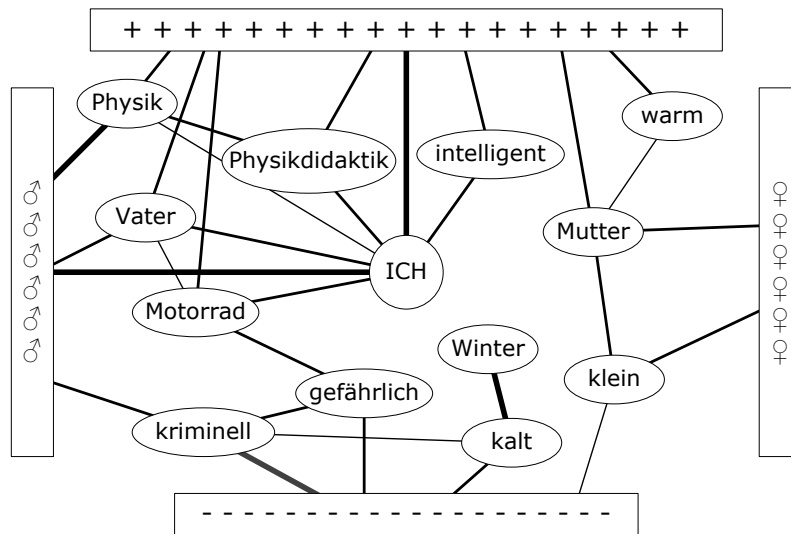


Abb. 5.1 Eine *Social Knowledge Structure* (SKS) für einen männlichen Physikdidaktiker. Die Konzepte werden durch Knoten (Ovale oder Rechtecke) dargestellt und die Assoziationen durch Verbindungslinien. Die Dicke der Verbindungslinie ist ein Maß für die Assoziationsstärke (angelehnt an Greenwald et al., 2002, S.5).

5.1 Definitionen, Prinzipien und empirische Überprüfung der BIT

Die BIT (Greenwald et al., 2002; Cvencek et al., 2012) verwendet zum Teil nicht streng definierte Begriffe, die allerdings sehr anschaulich sind und bereits in der Psychologie verwendet werden. Der Begriff *Konzept* steht für Personen, Gruppen oder Eigenschaften. Zwei spezielle Eigenschaftskonzepte sind die bewertenden Valenzkonzepte *positiv* und *negativ*. *Assoziationen* sind wechselseitige und moderierende Beziehungen zwischen zwei Konzepten. Die *Assoziationsstärke* ist die intervallskalierte Variable, die das Potenzial eines Konzepts darstellt, ein anderes Konzept zu aktivieren. Diese *Konzeptaktivierung* geschieht durch externe Stimuli oder eben durch Anregung des assoziierten und bereits aktivierten Konzepts.

Aufbauend auf diesen Begriffen beruht die BIT auf drei Vorannahmen: 1. Das soziale Wissen, d. h. das Wissen über Personen (auch das Selbst), über Gruppen und über deren Eigenschaften, kann als Netzwerk von Assoziationen mithilfe von Knoten (Konzepte) und Verbindungen (Assoziationen) dargestellt werden (*Social Knowledge Structure*, siehe Abb. 5.1). 2. Das Selbst ist zentrale Entität der assoziativen Wissensstruktur (SKS), was sich durch eine besonders starke Vernetzung innerhalb der Struktur deutlich macht. 3. Bewertende Eigenschaften können durch positive und negative Valenzknoten darge-

¹Die drei Theorien sind im Originalen veröffentlicht von Osgood und Tannenbaum (1955), Festinger (1957) und Heider (1958).

stellt werden und in den meisten Fällen ist das Selbst stärker mit *Positiv* als mit *Negativ* vernetzt.

Ein wesentlicher Bestandteil der Theorie sind die Definitionen, die bekannte kognitionspsychologische Konstrukte auf Assoziationen zurückführen. Dies wird letztendlich bei allen Schlussfolgerungen von IAT-Messungen auf diese Konstrukte ausgenutzt, da das IAT-Ergebnis eigentlich eine Operationalisierung von Assoziationen ist (siehe Abschnitt 4.1, S. 58, erster Absatz).

Definition (Einstellung): Eine Einstellung ist die Assoziation eines sozialen Objektes oder eines sozialen-Gruppen-Konzeptes mit einem Valenzkonzept.

In dem SKS des männlichen Physikdidaktikers (Abb. 5.1) ist z. B. die Verbindungen von *Mutter* zu *positiv* eine solche Einstellung bezüglich des Konzepts *Mutter*.

Definition (Stereotyp): Ein Stereotyp ist die Assoziation eines sozialen Objektes oder eines sozialen-Gruppen-Konzeptes mit einer nicht bewertenden Eigenschaft (kein Valenzkonzept).

So sind z. B. die Verbindungen zwischen *Physik* und *männlich* oder zwischen *Motorrad* und *gefährlich* Stereotypen. Eine Bewertung in Form einer Einstellung liegt im zweiten Fall zunächst nicht vor, auch wenn *gefährlich* wiederum mit *negativ* assoziiert wird.

Definition (Selbstwert): Der Selbstwert (auch das Selbstwertgefühl) ist die Assoziation des Konzeptes *Selbst* mit einem Valenzkonzept.

Als Selbstwert kann die Gesamtheit an direkten und indirekten Verbindungen zum Valenzkonzept aufgefasst werden. So tragen in Abb. 5.1 die Konzepte *Physikdidaktik* und *intelligent* zum positiven Selbstwert bei, indem sie als Mediatoren wirken (siehe später in Prinzip 1).

Definition (Selbstkonzept): Das Selbstkonzept ist die Assoziation des Konzeptes *Selbst* mit einer oder mehreren nicht bewertenden Eigenschaften (keine Valenzkonzepte).

Die Definition ist in dieser Form sehr streng in Greenwald et al. (2002) formuliert und berücksichtigt nur die Assoziationen vom *Selbst* zu Eigenschaftskonzepten. Demnach würden im obigen Beispiel nur die Verbindungen zu *intelligent* und *männlich* das Selbstkonzept darstellen. An anderer Stelle bemerken die Autoren jedoch, dass das Selbstkonzept auch die Verbindungen zu sozialen Kategorien wie z. B. *Vater* beinhaltet. In einer erweiterten Definition, in der Konzepte auch Schulfächer, Wissenschaftsgebiete oder sogar Fahrzeugkategorien vertreten können, würden zum Selbstkonzept auch deren Verbindungen zum Selbst zählen (siehe Erweiterung des Konzeptbegriffes in Abschnitt 5.2).

Die bisherigen Definitionen würden zunächst die BIT nur als beschreibendes Modell für Assoziationsstrukturen (SKS) nutzbar machen. Damit jedoch auch Prognosen und kausale Aussagen aus der Theorie heraus getroffen werden können, sind die drei folgenden Definitionen und Prinzipien notwendig.

Definition (Shared first-order link): Wenn zwei Knoten mit demselben dritten Knoten verbunden sind, teilen sie sich eine Verbindung erster Ordnung.

Prinzip 1 (Balance-congruity): Wenn sich zwei unverbundene oder schwach verbundene Knoten eine Verbindung erster Ordnung teilen, wird sich die Assoziation zwischen diesen beiden verstärken.

Die beiden Konzepte *klein* und *weiblich* teilen sich eine Verbindung erster Ordnung, indem sie beide direkt mit *Mutter* verbunden sind. Dies hat zu Folge, dass sich dadurch auch die Assoziation zwischen *klein* und *weiblich* verstärken wird.

Definition (Bipolar opposition of nodes): Wenn sich zwei Knoten weniger Verbindungen erster Ordnung teilen, als durch Zufall erwartet, können sie als bipolar-gegensätzlich bezeichnet werden.

Prinzip 2 (Imbalance-dissonance): Das SKS-Netzwerk widerstrebt der Bildung von neuen Verbindungen, die als Resultat einen Knoten mit Verbindungen erster Ordnung zu bipolar-gegensätzlichen Knoten hätten.

Bipolar-gegensätzliche Knoten sind z. b. die Konzepte *positiv* und *negativ*, *männlich* und *weiblich*. In dem Beispiel in Abb. 5.1 könnten jedoch auch die Konzepte *warm* und *kalt* als bipolar-gegensätzlich bezeichnet werden. Letztendlich trifft dies auf alle Konzeptpaare zu, die sich auch als Kategorienpaar eines IATs eignen würden (siehe Abschnitt 4.1.1). Für die Darstellung im SKS gibt es keine spezielle Regel, wann für ein Knoten ein Oval oder ein Rechteck verwendet wird. Allerdings eignen sich wegen der Übersichtlichkeit besonders bipolar-gegensätzliche Knoten als Rechtecke am Rand mit dem Selbst im Zentrum.

Prinzip 2 ist das notwendige Gegenstück zu Prinzip 1, denn es verhindert die inflationäre Bildung von Assoziationen jeder Art, die bei jeder Verbindung erster Ordnung auftreten würden. So wehrt sich das SKS-Netzwerk trotz der geteilten Verbindung erster Ordnung gegen eine Assoziation zwischen *Mutter* und *negativ*, weil das wiederum eine Verbindung erster Ordnung zwischen *positiv* und *negativ* bedeuten würde.

Definition (Pressured concept): Ein Konzept ist gedrängt, wenn es als Folge anhaltender oder wiederholender Einflüsse (durch das balance-congruity-Prinzip) Verbindungen zu bipolar-gegensätzlichen Knoten bilden würde.

Prinzip 3 (Differentiation): Gedrängt Konzepte neigen dazu, sich in Subkonzepte aufzuteilen, die mit jeweils einem der beiden drängenden bipolar-gegensätzlichen Knoten verbunden sind.

In Abb. 5.1 könnte *Motorrad* zu einem gedrängten Konzept werden, wenn sich z. b. durch viele Zeitungsmeldungen über Motorradunfälle die Assoziation zu *gefährlich* verstärken würde und somit durch Prinzip 1 (balance-congruity) auch die Assoziation zu *negativ*. Der Druck des gedrängten Konzepts wird durch Prinzip 3 abgebaut, indem *Motorrad* in zwei Subkonzepte aufgeteilt wird, in diesem Fall positive und negative Aspekte des Motorrads bzw. des Motorradfahrens (z. b. Fahrspaß, Beschleunigung und Ausflüge vs. Gefahren, Pflege und Kosten).

Die hier vorgestellten Prinzipien verdeutlichen den Charakter der BIT als vereinheitlichende Theorie. Schon die Bezeichnungen der Prinzipien machen ihre Herkunft aus der Balance-, Dissonanz- und Kongruitätstheorie sichtbar. Die BIT basiert somit auf einer langjährigen Forschungstradition. Dabei begreift sie jedoch konsequent die oben genannten Konstrukte als Assoziationen, die prinzipiell immer mit einem IAT messbar sind.

Die enge Verknüpfung zwischen Balanced Identity Theory und impliziten Assoziationstests zeigt sich vor allem in den Methoden, die für die empirische Fundierung der Theorie verwendet werden. Während andere Konsistenztheorien vor allem mit expliziten Messungen erforscht wurden, werden bei der BIT explizite und implizite Verfahren eingesetzt, wobei die impliziten Messungen die Vorhersagen der Theorie besser bestätigen (Cvencek et al., 2012).

Das von Greenwald et al. (2002) vorgeschlagene Balanced Identity Design (BID) zur Überprüfung von Prinzip 1 untersucht immer gleichzeitig die Abhängigkeit von drei verschiedenen Assoziationen untereinander, eine Gruppen-Selbst-Assoziation (Identität), eine Selbst-Eigenschaft-Assoziation (Selbstkonzept oder Selbstwert) und eine Gruppen-Einstellung-Assoziation (Stereotyp oder Einstellung, siehe Abb. 5.2). Im Beispiel in Abb. 5.1 wären die bestehenden Assoziationen des Selbst zur Gruppe der Männer und zur positiven Valenz der untere Teil des Dreiecks in Abb. 5.2. *Positiv* und *Männer* teilen sich somit eine Verbindung erster Ordnung. Prinzip 1 sagt nun voraus, dass sich damit die Assoziation von *Männer* und *positiv* verstärkt. Das BID sähe nun die implizite und explizite Messung der drei Assoziationen vor, um anschließend mithilfe multipler Regressionen zu untersuchen, inwiefern jeweils zwei der Assoziationen die Stärke der Dritten vorhersagen. Die genaue Beschreibung des vierschrittigen Verfahrens findet man bei Greenwald et al. (2002).

Das bisherige Forschungsinteresse vergangener Studien lag laut Cvencek et al. (2012) lediglich in der empirischen Überprüfung von Prinzip 1 (Balance-congruity), weshalb sich die Autoren in ihrer Meta-Analyse auch auf diesen Aspekt beschränken. Letztendlich werden von ihnen 17 Veröffentlichungen mit 18 unabhängigen Stichproben ($N \approx 1900$) berücksichtigt. Alle ausgewählten Studien beinhalten Messungen von drei Assoziationen², auf die das BID angewendet werden kann. Zusammengefasst ist das Ergebnis der Meta-Analyse, dass die implizit gemessenen Assoziationen die Vorhersagen aus

²18 Stichproben mit jeweils 3 impliziten Messungen und 8 Stichproben mit jeweils 3 expliziten Messungen

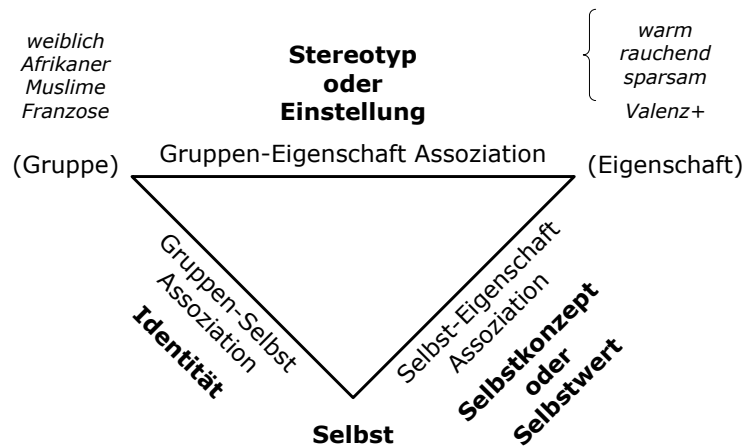


Abb. 5.2 Darstellung des Balanced Identity Designs (BID). Jede Ecke stellt ein Konzept dar. Das BID beinhaltet immer das *Selbst* als eines der Konzepte, hinzu kommt ein Gruppen- und ein Eigenschaftskonzept. Die *kursiv* gedruckten Gruppen oder Eigenschaften sind Beispiele für Konzepte, die mit dem BID untersucht werden könnten. Die drei im BID gemessenen Assoziationen stellen die Kanten des Dreiecks dar. Sie können je nach Konzeptpaar eine Identität, ein Stereotyp, eine Einstellung, ein Selbstkonzept oder der Selbstwert sein (übersetzt nach Greenwald et al., 2002, S.9).

Prinzip 1 signifikant besser bestätigen ($Mn .27 \leq r \leq .35$, mittlere Effekte), als die expliziten Messungen ($Mn .08 \leq r \leq .18$, kleine Effekte). Damit ist ein empirischer Beleg dafür geliefert, dass die Messungen mit einem IAT sich gut durch die BIT beschreiben lassen bzw. einige Vorhersagen aus der Theorie mit den Messungen des IATs korrespondieren.

5.2 Balanced Identity und das Image von Physik

Während in Abschnitt 5.1 das *Konzept* für Personen, Gruppen, oder deren Eigenschaften steht (vgl. Greenwald et al., 2002), also eher aus sozialpsychologischer Perspektive definiert wird, ist es für die Anwendung der BIT auf eine physikdidaktische Fragestellung sinnvoll, als *Konzepte* auch weitere Kategorien zuzulassen. Bei Cvencek et al. (2012) sind bereits die Knoten *science*, *math*, *arts* und *poetry* vorhanden, was nahelegt, dass auch Wissenschaftsdomänen eine SKS bereichern können. Dafür spricht auch ihre Anwendung des Balanced Identity Designs auf das Konzept *school*, was als Gruppenkonzept interpretiert wird (siehe Tabelle 8.1. in Cvencek et al., 2012, S.165). Wenn als Konzepte nur Personen, Gruppen und deren Eigenschaften infrage kommen sollen, könnte auf diesem Wege das Schulfach *Physik* trotzdem verwendet werden, indem es als die Gruppe der Prototypen zum Fach interpretiert wird.

Aber auch ein erweiterter Blick auf den Begriff des *Selbstkonzepts* zieht andere sinnvolle Kategorien nach sich, die als *Konzept* bzw. Knoten infrage kommen. In einer frühen

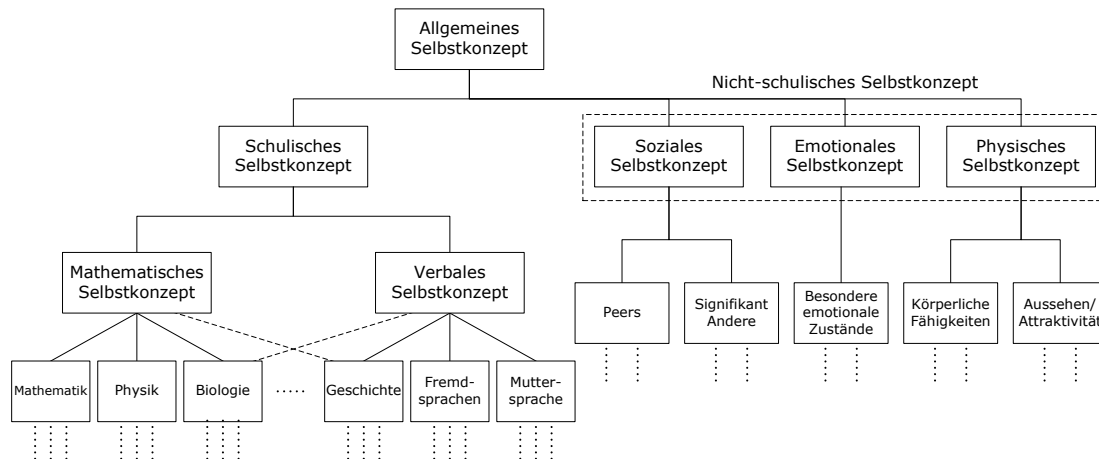


Abb. 5.3 Mehrdimensionales Selbstkonzept (modifiziert und übersetzt nach Shavelson et al., 1976; Marsh, 1990, S. 413 und 93). Die ursprünglich hierarchische Struktur konnte innerhalb des schulischen Selbstkonzeptes empirisch nicht bestätigt werden, sodass von den zwei übergeordneten Faktoren *mathematisches* und *verbales Selbstkonzept* auszugehen ist.

Modellierung des Selbstkonzepts unterscheiden Shavelson, Hubner und Stanton (1976) bereits auf der zweiten Hierarchieebene zwischen einem schulischen und einem nicht-schulischen Selbstkonzept (siehe Abbildung 5.3). Auch wenn die hierarchische Struktur empirisch nicht bestätigt wird, so kann jedoch sehr wohl gezeigt werden, dass das Selbstkonzept ein mehrfaktorielles Konstrukt ist (vgl. Möller & Trautwein, 2009). Die in Abschnitt 5.1 vorgenommene Definition des Selbstkonzepts als Assoziation des Selbst mit nicht bewertenden Eigenschaften wäre in diesem Modell durch das emotionale und das physische Selbstkonzept vertreten. Die im BID benötigte Assoziation des Selbst mit einer Gruppe (*Identität*, siehe Abb. 5.2) wäre hier durch das Soziale Selbstkonzept ebenfalls integriert. Hinzu kommt außerdem das schulische Selbstkonzept, welches bisher in der BIT nicht berücksichtigt ist. Das könnte mit dem Umstand begründet werden, dass das schulische Selbstkonzept schwerpunktmäßig in der pädagogischen Psychologie erforscht wird, während die BIT ihre Wurzeln in der Sozial- und Kognitionspsychologie hat.

Bei einem fachdidaktischen Hintergrund ist es also zweckmäßig, die BIT auf ein größeres Feld anzuwenden, indem als Konzepte auch Schulfächer oder Wissenschaftsdomänen zugelassen werden. Außerdem sollte das Selbstkonzept eher an das in der pädagogischen Psychologie verwendete Konstrukt angelehnt sein, um die speziell im schulischen Kontext bestehenden Fähigkeitsselbstkonzepte abbilden zu können. Somit werden im Folgenden unter *Selbstkonzept* alle Assoziationen des Selbst mit einer oder mehreren (nicht bewertenden) Eigenschaften, Gruppen, Schulfächern oder Wissenschaftsdomänen verstanden. In dem Beispiel in Abb. 5.1 konstruiert sich das Selbstkonzept demnach aus den Verbindungen des *Ich-Knotens* zu *Vater*, *Physikdidaktik* und *Physik*. Die Verbindung zu *Motorrad* könnte in einer weitergehenden Verallgemeinerung der Definition ebenfalls

zum Selbstkonzept zählen. Da dies aber für die Fragestellung der vorliegenden Arbeit nicht relevant ist, genügt es, sich auf die hier vorgenommene Erweiterung der BIT zu beschränken.

Die hier vorgenommene nähere Auseinandersetzung mit dem Selbstkonzept im Rahmen der BIT ist deshalb wichtig, weil durch einen physikdidaktischen Ansatz zunächst nur angestrebt werden kann, das fachbezogene Selbstkonzept oder die fachbezogenen Stereotypen und Einstellungen (siehe Definitionen in Abschnitt 5.1) positiv zu beeinflussen. Ein positiver Einfluss auf den Selbstwert wäre zwar nicht von Nachteil, ist jedoch eher ein allgemein pädagogisches Ziel. Der positive Einfluss eines hohen fachbezogenen Selbstkonzepts auf die Kompetenzentwicklung und das Interesse zum Fach gilt als empirisch gesichert (Möller & Trautwein, 2009). Während das Selbstkonzept und die Leistung in einem sich gegenseitig verstärkenden Zusammenhang stehen³, fördert das Selbstkonzept auch die Interessenentwicklung. Diese wiederum steht ebenfalls im positiven Zusammenhang mit lernförderlichen Verhaltensweisen, sodass die Autoren als Fazit festhalten: „Zusammengefasst kann die Förderung eines adäquaten und positiven Selbstkonzepts als zentrales Ziel pädagogischer Bemühungen gelten.“ (Möller und Trautwein, 2009, S. 201)

In Abb. 5.4a und 5.4b sind zwei assoziative Wissensstrukturen (SKSs) skizziert, die denkbare Extremfälle für das physikbezogene Selbstkonzept darstellen. *Physik* wird im Mittel eher mit *männlich* und *schwierig* assoziiert (Kessels et al., 2006), was laut der Definition zwei Stereotypisierungen sind. Daher gibt es in beiden Fällen die Verbindungen vom *Physik*-Knoten zu den beiden Eigenschaftsknoten. Im ersten Extrembeispiel (Abb. 5.4a) steht ein leistungsmotivierter Junge im Zentrum, der sein Selbst eher mit *schwierig* und *männlich* assoziiert. Die Folge von Prinzip 1 wäre eine Verstärkung der Verbindung zwischen dem Selbst und *Physik* (Selbstkonzept), da sich die beiden Konzepte zweimal eine Verbindung erster Ordnung teilen, einmal bei *männlich* und einmal bei *schwierig*. Im zweiten Extrembeispiel (Abb. 5.4b) steht ein misserfolgsmeidendes Mädchen im Zentrum, mit dementsprechenden Verbindungen des Selbst zu *leicht* und *weiblich*. Eine Verbindung zwischen dem Selbst und *Physik* kann in diesem Fall nur mit starkem Widerstand entstehen, da nach Prinzip 1 daraus auch Verbindungen zwischen dem Selbst und *männlich* und *schwierig* gefördert werden würden. Nach Prinzip 2 widersetzt sich jedoch das Netzwerk gegen diese Verbindungen, da dann Verbindungen erster Ordnung zu bipolar gegensätzlichen Knoten entstehen würden. Würde durch äußere Einflüsse, z. B. ein speziell an den Interessen von Mädchen orientierter Physikunterricht, dennoch Assoziationen zwischen dem Selbst und *Physik* entstehen, wäre das Resultat ein gedrängtes *Ich*-Konzept. Mit anderen Worten: Wenn die Stereotypen zu *Physik* zum Selbstkonzept passen, wird dadurch das fachbezogene Selbstkonzept gefördert. Sind jedoch die Stereotypen nicht passend, behindern die Prinzipien der SKS eine Identifizierung mit dem Fach bzw. das Entstehen eines positiven physikbezogenen Selbstkonzepts.

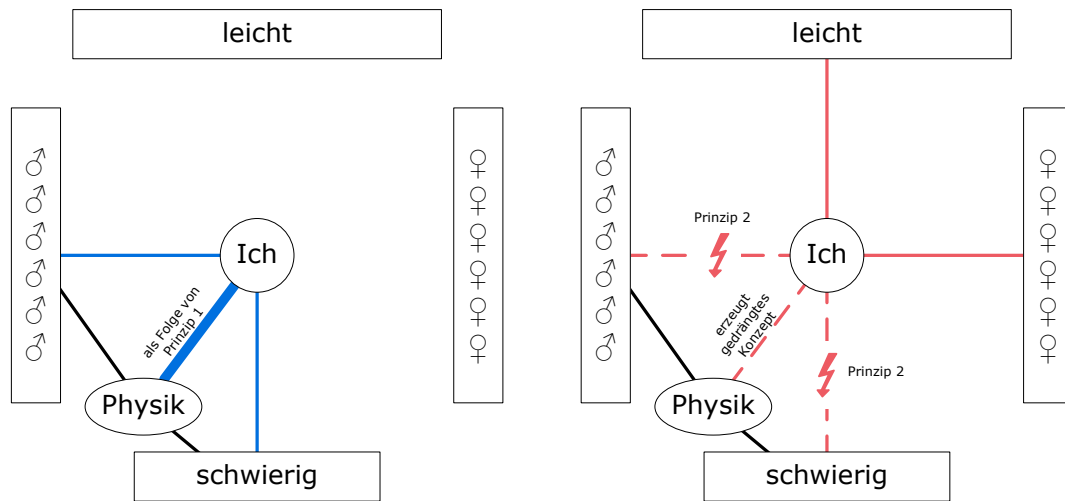
³Gute Leistungen fördern wiederum das Selbstkonzept.

Der oben erwähnte physikdidaktische Ansatz kann nur darin bestehen, auf die Stereotypen des *Physik*-Konzeptes Einfluss zu nehmen und damit das physikbezogene Selbstkonzept zu fördern. Denn die Auswirkungen des Physikunterrichts auf den Selbstwert oder das allgemeine Selbstkonzept der Schülerinnen und Schüler kann als gering eingeschätzt werden. Zudem ist eine derartige Einflussnahme auch nicht das Anliegen physikdidaktischer Forschung. Als ein ideales Ziel könnte der Zustand aufgefasst werden, in dem *Physik* unabhängig von bestimmten Konzepten ist. In dem vorliegenden Fall wäre es von Vorteil, wenn die beiden Stereotypen („Physik ist männlich und schwierig“) gar nicht existieren würden oder nur so schwach wären, dass sich im zweiten Extrembeispiel ein positives physikbezogenes Selbstkonzept entwickeln kann, ohne dass ein gedrängtes Konzept entsteht. Der leistungsmotivierte Junge (Abb. 5.4c) profitiert im Vergleich zu der oben beschriebenen Situation (Abb. 5.4a) nicht von diesem Zielzustand. Dennoch wird die Bildung einer Verbindung zu *Physik* (physikbezogenes Selbstkonzept) nicht behindert. Im Fall des misserfolgsmeidenen Mädchens (Abb. 5.4d) ist nun auch ein physikbezogenes Selbstkonzept ohne Widerstand möglich. Zudem wären in diesem individuellen Netzwerk dann auch Verbindungen von *Physik* zu *leicht* und *weiblich* möglich, weil sich die beiden Konzepte mit *Physik* Verbindungen erster Ordnung teilen (Prinzip 1). Durch die vorab fehlenden Stereotypen des *Physik*-Konzeptes wäre die Entwicklung des physikbezogenen Selbstkonzepts in beiden Fällen unabhängig vom allgemeinen Selbstkonzept des jeweiligen Individuums.

5.3 Bedeutung für die vorliegende Arbeit

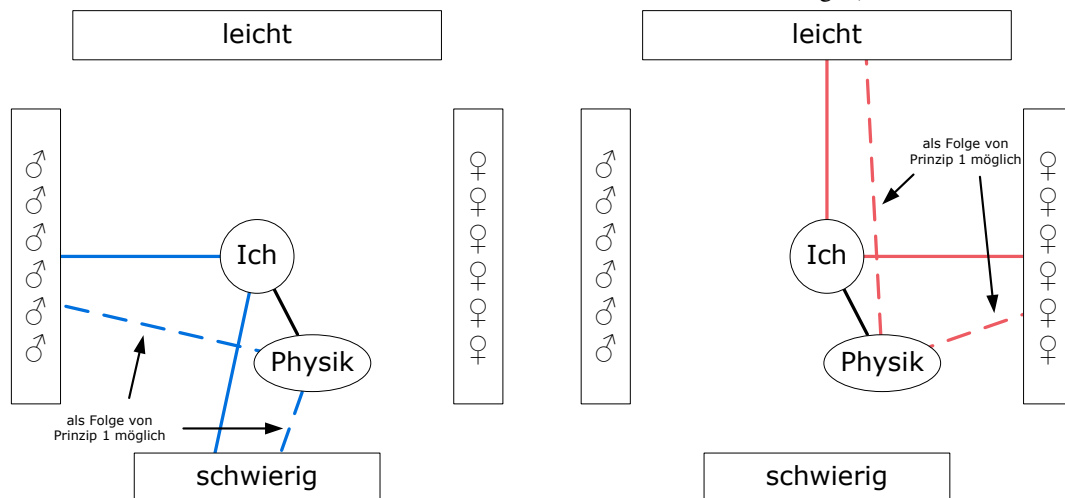
Mithilfe der Balanced Identity Theorie können die mit dem IAT ermittelten empirischen Befunde erfolgreich beschrieben und erklärt werden. Durch sie werden die im Rahmen dieser Arbeit relevanten Konstrukte *Stereotyp* und *Selbstkonzept* als Assoziationen interpretiert, die somit durch einen IAT implizit zugänglich sind. Damit ist der in Kapitel 3 formulierte Zusammenhang zwischen dem vermittelnden Charakter PbPUs und dem fachbezogenen Selbstkonzept auch durch IATs direkt prüfbar. Nach den Prinzipien der BIT würde solch ein lernförderliches Selbstkonzept durch die Veränderung des Images von Physik gefördert werden, indem die Stereotypisierungen zu *schwierig* und *männlich* aufgelöst werden. Falls PbPU auf diese Stereotypisierung jedoch keinen Einfluss hat und sich aufgrund seiner Merkmale dennoch positiv auf das fachbezogene Selbstkonzept auswirkt, würde das einer Differentiation des Konzepts *Physik* entsprechen (Prinzip 3).

Die Balanced Identity Theorie stellt in der vorliegenden Arbeit den theoretischen Rahmen für Erklärungen und Schlüsse aus den Ergebnissen der IATs dar. Sie ermöglicht Interpretationsmöglichkeiten und Vorhersagen über die assoziativen Zusammenhänge des Selbstkonzepts und des Fachimages.



(a) SKS eines leistungsmotivierten Jungen (blaue Verbindungen). Physik wird im Mittel eher mit *schwierig* und *männlich* assoziiert (schwarze Verbindungen). Nach Prinzip 1 verstärkt sich durch die beiden geteilten Verbindungen erster Ordnung die Assoziation zu *Physik* (dicke blaue Verbindung).

(b) SKS eines misserfolgsmeidenden Mädchens (rote Verbindungen). Physik wird im Mittel eher mit *schwierig* und *männlich* assoziiert (schwarze Verbindungen). Das Netzwerk widerstrebt nach Prinzip 2 der Verbindung zwischen dem Selbst und *Physik*, weil sie geteilte Verbindungen erster Ordnung zu bipolar gegensätzlichen Konzepten erzeugen würde (rot gestrichelte Verbindungen).



(c) Auswirkung eines Physik-Konzeptes ohne von vornherein bestehenden Stereotypen auf das SKS eines leistungsmotivierten Jungen. Eine Assoziation zu *Physik* ist möglich (schwarze Verbindung) und würde die Entstehung der individuellen Stereotypen fördern (blau gestrichelte Verbindungen).

(d) Auswirkung eines Physik-Konzeptes ohne von vornherein bestehenden Stereotypen auf das SKS eines misserfolgsmeidenden Mädchens. Eine Assoziation zu *Physik* ist möglich (schwarze Verbindung) und würde Assoziationen von *Physik* zu *leicht* und *weiblich* fördern (Stereotypen, rot gestrichelte Verbindungen).

Abb. 5.4 Ausgangslage und Ziel für das Image von Physik und dessen Auswirkungen auf die assoziativen Wissensstrukturen von zwei Extrembeispielen.

Forschungsfragen und Hypothesen

6.1 Entwicklung der Forschungsfragen

In dem bisherigen Teil der Arbeit wurde phänomenbasierter Physikunterricht als ein curricularer Ansatz vorgestellt, der didaktisch, fachlich und wissenschaftstheoretisch fundiert und darüber hinaus langjährig erprobt ist. Für den Inhaltsbereich der Anfangsoptik ist er bereits soweit etabliert, dass er in Rahmenlehrplänen und Schulbüchern berücksichtigt wird. Trotz dieser Gegebenheit gibt es bisher kaum empirische Untersuchungen zur Wirkungsweise des PbPUs auf die Interessen- und Leistungsentwicklung von Schülerinnen und Schülern. Dadurch entzieht er sich der Vergleichbarkeit mit anderen curricularen Ansätzen, die eine empirische Fundierung vorweisen können. Eine solche Fundierung für die verschiedenen Aspekte des PbPUs zu erbringen, stellt eines der dringlichsten Forschungsdesiderate für die phänomenbasiert orientierte Physikdidaktik dar. Daraus leitet sich das Motiv der vorliegenden Arbeit in Form einer allgemeinen Forschungsfrage ab:

Wie wirkt sich phänomenbasierter Unterricht auf das Lernen von Physik aus?

In den vorangegangenen Kapiteln wurde der Zusammenhang zwischen der Zielsetzung PbPUs, das Interesse für den Unterricht und dessen Inhalte zu fördern, dem Image des Faches, dem fachbezogenen Selbstkonzept und den Assoziationen zur Physik dargestellt (zusammenfassend in den Abschnitten 2.6, 3.4, 4.3 und 5.3). Im Rahmen dieser Arbeit soll ein Teil dazu beigetragen werden, dem Mangel an empirischer Fundierung zu begegnen, indem mit der Anfangsoptik ein repräsentatives Beispiel für PbPU darauf

hin untersucht wird. Mit dem Nachweis wäre eine weitere Voraussetzung für erfolgreiches Lernen von PbPU gewährleistet, die sich nicht unmittelbar aus seinen Merkmalen ableiten lässt (siehe Abschnitt 2.3). Daraus ergeben sich spezifisch für diese Arbeit insgesamt vier Forschungsfragen, deren Antworten Beiträge für die oben allgemein formulierte Fragestellung liefern sollen. Die Erste von ihnen ist ursächlich für die folgenden.

Forschungsfrage 1 (Wirkung auf das Interesse): Welchen Einfluss hat phänomenbasierter Physikunterricht der Anfangsoptik auf das Interesse der Schülerinnen und Schüler an Physik?

In Kapitel 3 wurde dargelegt, dass es für eine solche Einflussnahme notwendig ist, das Image des Faches zu verbessern, da es sonst im Konflikt mit der Selbstkonzeptentwicklung der Schülerinnen und Schüler stehen kann. In Abschnitt 5.2 wurde erläutert, dass dazu die bestehenden Assoziationen (Stereotypen) zu *schwierig* und *männlich* aufgelöst werden müssten. Das führt zu der zweiten Forschungsfrage, deren Antwort somit gewissermaßen eine Bedingung für die der ersten Frage ist.

Forschungsfrage 2 (Wirkung auf das Image): Welchen Einfluss hat phänomenbasierter Physikunterricht der Anfangsoptik auf das Image von Physik als schwieriges und männliches Fach?

In Abschnitt 5.2 wurde außerdem begründet, warum eine Veränderung des Images zu einem bezüglich der Schwierigkeit und Geschlechtskonnotation neutralem Fach, die Entstehung eines positiven fachbezogenen Selbstkonzepts fördern würde. Dies wiederum ist das Bindeglied zwischen dem in Abschnitt 3.2 dargestellten Zusammenhang zwischen Image und Interesse. Die dritte Forschungsfrage nimmt somit eine vermittelnde Rolle zwischen der ersten und der zweiten Frage ein.

Forschungsfrage 3 (Wirkung auf das Selbstkonzept): Welchen Einfluss hat phänomenbasierter Physikunterricht der Anfangsoptik auf das physikbezogene Selbstkonzept der Schülerinnen und Schüler?

Einen bisher wenig diskutierter Aspekt, der jedoch in einer bewertenden Untersuchung von Curricula nicht vernachlässigt werden darf, stellen die im Unterricht erworbenen Kompetenzen dar, z. B. in Form von Fachwissen. Die Berücksichtigung des Fachwissens hat im Wesentlichen zwei Gründe. Das sind zunächst die in den Abschnitten 2.3 und 3.1 dargestellten positiven Zusammenhänge zwischen den hier zu überprüfenden motivationalen Konstrukten und erfolgreichem Lernen bzw. der Leistung im Fach. Demnach könnte phänomenbasierter Physikunterricht, der sich auf das Interesse, das Image und das Selbstkonzept positiv auswirkt, auch die Konstruktion des Fachwissens in diese Richtung beeinflussen. Andererseits beziehen sich diese positiven Zusammenhänge zumeist nur auf die langfristige *Qualität des Wissens*. Der zweite Grund liegt darin,

dass kontrolliert werden muss, ob eine alternative Unterrichtskonzeption wie PbPU sich zumindest nicht negativ auf die erworbenen Kompetenzen auswirkt. Dies könnte der Fall sein, da PbPU aufgrund seiner Merkmale und Zielsetzungen eine Verschiebung der Schwerpunkte von Unterrichtsinhalten und -tätigkeiten vorsieht, indem er z. B. Raum für affektive Aspekte und explorative Phasen schafft oder auf sonst verwendete Modellvorstellungen verzichtet. Das führt zu der vierten und letzten Forschungsfrage, die eher Kontrollzwecken dient und kein vorrangiges Forschungsinteresse darstellt.

Forschungsfrage 4 (Wirkung auf das Fachwissen): Welchen Einfluss hat phänomenbasierter Physikunterricht der Anfangsoptik auf das Fachwissen der Schülerinnen und Schüler?

6.2 Hypothesen

An dieser Stelle sollen die zu den Forschungsfragen aus Abschnitt 6.1 gehörigen *inhaltlichen* Hypothesen begründet und aufgeführt werden. Die auf das Untersuchungsdesign und Auswerteverfahren angepassten *statistischen* Hypothesen werden in Abschnitt 7.3 formuliert. Die Begründung für die auf Forschungsfrage 1 bezogene Hypothese lässt sich aus Hypothese 2 und 3 ableiten, weshalb sie nach ihnen erfolgt. Die Nummerierung erfolgt hier deshalb nicht fortlaufend, damit die Zuordnung von Hypothese zu Forschungsfrage erhalten bleibt.

Für die Begründung einer Hypothese bezüglich des Images von Physik können allein mithilfe der IAT-Ergebnisse von Kessels et al. (2006) nur schwer Aussagen getroffen werden, da die Mechanismen nicht bekannt sind, die z. B. zu einer Assoziation von *Physik* mit *schwierig* geführt haben. Nach den Prinzipien der BIT könnte z. B. eine starke Assoziation von *Mathematik* mit *schwierig* bestehen. Eine gleichzeitig bestehende Assoziation von *Mathematik* mit *Physik* wäre eine geteilte Verbindung erster Ordnung, die zu dem Image von Physik als schwieriges Fach führt. Diese Vermutung wird gestützt durch die Tatsache, dass Mathematik das einzige Fach ist, welches als noch schwieriger betrachtet wird, als Physik (Watson et al., 1994, siehe Abschnitt 3.2). Jedoch sind die Annahmen über die genannten Assoziationen nur spekulativ und müssten überprüft werden, was von der Fragestellung dieser Arbeit zu weit entfernt wäre.

Mehr Aufschluss geben jedoch die Items von Watson et al. (1994), die überhaupt erst zu der Bewertung von Physik und Mathematik als schwierige Fächer geführt haben. Die Fächer sind als schwierig bezeichnet worden, weil sie das Verstehen abstrakter Prinzipien beinhalten, weil die Gedankengänge oft verwirrend sind, weil logisches Denken gefordert ist, weil ein hohes Maß an Problemlösekompetenz gefordert ist und weil für das Fach ein natürliches Talent nötig ist. Prinzipiell sind das die Kompetenzen, die durch den Physik-, vor allem aber durch den Mathematikunterricht gefördert werden sollen. Bis auf die Punkte, dass das Fach teilweise verwirrend ist und ein besonderes Talent erfordert, wäre eine derartige Beurteilung von Physikunterricht demnach kein Makel.

Trotzdem kann bereits hier argumentiert werden, dass PbPU vom Ansatz her weiter von der Mathematik entfernt ist, als das modellbasierte Vorgehen. Die Ausgangs- und Endpunkte sind die Phänomene und die Naturbeobachtung und nicht die Modelle, aus denen die Beobachtungen deduziert werden. Dies würde eher dem strengen mathematischen Umgang mit Theorien entsprechen. PbPU hebt also gerade das vom Physikunterricht hervor, was ihn vom Mathematikunterricht unterscheidet: der Bezug zur wahrgenommenen Realität und die Aufgabe, diese zu beschreiben.

Bezüglich der genannten Eigenschaften, die der Physikunterricht laut den Ergebnissen von Watson et al. (1994) hat und auch haben soll, ist es also wichtig, die *subjektiv wahrgenommene Schwierigkeit* zu betrachten, da diese nicht im Konflikt mit dem Selbstkonzept stehen darf. Das Fach darf also durchaus als *objektiv anspruchsvoll* von jedem wahrgenommen werden. Es muss jedoch für jeden nachvollziehbar und mit Bezug zum Selbst und zur eigenen Lebenswelt sein.

Genau diese Punkte werden durch PbPU gewährleistet. Das eigene Erleben, die Subjektivität und die Einbindung machen das sonst zu Abstrakte und Verwirrende von jedem nachvollziehbar. Der Unterricht wird nicht als verwirrend wahrgenommen, weil er nicht den selbst gemachten Erfahrungen widerspricht, sondern auf sie aufbaut. Das Verstehen beruht auf eigens konstruierte Sinnzusammenhänge und der kritische Umgang mit Modellen führt nicht zu ontologischen Missverständnissen, wodurch der Umgang mit widersprüchlichen Modellen ebenfalls als schwierig aufgefasst werden könnte.

Auch dem zweiten Gesichtspunkt des Images von Physik, die Assoziation mit *männlich* wird durch PbPU begegnet. Der von Häußler et al. (1996) identifizierte Interessentyp A unterscheidet sich von den anderen Typen durch sein Interesse an der Berechnung physikalischer Größen, der quantitativen Seite der Physik, der Planung von Experimenten, an Einblicke in technische Berufe und an die Funktionsweise von Geräten. Dieser Typ A hat in Physik überwiegend gute und mittlere Noten und besteht zu 80 % aus Jungen. Er stellt somit den Physik-Prototypen dar und ist folglich der Grund für die Assoziation mit *männlich* (siehe Abschnitt 3.2). Aus seinen Merkmalen geht hervor, dass PbPU gerade diese Stereotypen nicht bedient. Damit wird Typ A nicht zum Physik-Prototyp und das Image des Faches eher geschlechtsneutral.

Hypothese 2: Phänomenbasierter Physikunterricht der Anfangsoptik wirkt sich positiv auf das Image von Physik aus.

H 2.a: Phänomenbasierter Physikunterricht der Anfangsoptik schwächt die Assoziation von *Physik* mit *schwierig*.

H 2.b: Phänomenbasierter Physikunterricht der Anfangsoptik schwächt die Assoziation von *Physik* mit *männlich*.

Bei einer Verbesserung des Images von Physik durch die Abschwächung der Assoziationen zu *schwierig* und *männlich* lässt sich aus der BIT schlussfolgern, dass ein nicht gedrängtes physikbezogenes Selbstkonzept entstehen kann (siehe Abb. 5.4). Zusätzlich

tragen die Merkmale PbPUs (Aspekt der Mediation) zu einer Identifizierung mit dem Unterrichtsfach und seinen Inhalten bei (siehe S. 41).

Hypothese 3: Phänomenbasierter Physikunterricht der Anfangsoptik wirkt sich positiv auf das physikbezogene Selbstkonzept aus. Er stärkt die Assoziation zwischen dem *Selbst* und *Physik*.

Mit einer Verbesserung des Fachimages und der Förderung des physikbezogenen Selbstkonzepts sind die Voraussetzungen für eine positive Interessenentwicklung geschaffen. Hinzu kommt die Steigerung der gefühls- und wertbezogenen Bedeutsamkeit des Lerngegenstandes durch die Aspekte der Subjektivität, Mediation und Affektivität, die sich ebenfalls auf das Interesse positiv auswirken (siehe Definition Interesse, S. 42)

Hypothese 1: Phänomenbasierter Physikunterricht der Anfangsoptik erhöht das Interesse am Physikunterricht und seinen Inhalten.

In Bezug auf Forschungsfrage 4 lassen sich m. E. theoretisch keine Argumente dafür herleiten, dass ein phänomenbasiertes Vorgehen Nachteile im Fachwissen zur Folge hat. Die bisherigen Erfahrungen mit der Anfangsoptik, die empirisch nicht gesichert sind, sprechen eher für ein gutes Verständnis der Unterrichtsinhalte. Inwieweit PbPU spezifische Kompetenzen fördert oder vernachlässigt bleibt zu prüfen. Dies kann jedoch nicht im Rahmen der vorliegenden Arbeit geschehen.

Hypothese 4: Phänomenbasierter Physikunterricht der Anfangsoptik wirkt sich weder positiv noch negativ auf den Lernerfolg in Form von Fachwissen aus.

Teil II

Empirischer Teil

7.1 Wahl des Forschungsdesigns und der Analysemethode

Die Formulierung der allgemeinen Forschungsfrage in Kapitel 6 zielt auf einen *Vergleich* zwischen einem phänomenbasierten Curriculum und einem „Normalunterricht“, der bisher nicht genauer gekennzeichnet wurde. Das Interesse der vorliegenden Arbeit liegt darin, die Wirksamkeit PbPUs zu untersuchen, ohne ihm kontrastierend einen bestimmten Unterricht entgegenzustellen. Ein solcher Gegenpart könnte z. B. als modellbasierter Physikunterricht bezeichnet werden. Er müsste jedoch genauso wie PbPU anhand von Merkmalen charakterisiert werden und es müsste mit der gleichen Sorgfalt eine modellbasierte Unterrichtssequenz zur Anfangsoptik entworfen werden. Das würde jedoch die Objektivität der Studie maßgeblich beeinflussen, da so nicht sichergestellt wäre, dass ein fairer Vergleich stattfindet. Der hier getroffenen Entscheidung, den Vergleichsunterricht nicht näher festzulegen, liegt die Annahme zugrunde, dass sich der Unterricht von Lehrkräften, die bisher nicht mit PbPU in Kontakt gekommen sind, in hinreichend vielen Merkmalen von diesem unterscheidet. Die Forderungen an den Vergleichsunterricht lauten somit nur, dass er sich nicht in den Unterrichtsinhalten¹ und in der Dauer von der phänomenbasierten Anfangsoptik unterscheidet.

Mit der Festlegung auf einen derartigen Vergleich ist auch determiniert, dass es sich um eine *quasiexperimentelle Feldstudie* handeln muss. Da es sich bei dem Untersuchungsgegenstand um Unterricht handelt, der für den Schulalltag konzipiert wurde, soll er sich auch in diesem Umfeld bewähren. Eine Untersuchung unter Laborbedingungen wäre nur schwer generalisierbar und eventuell gefundene Einflüsse auf lernwirksame

¹Senatsverwaltung für Bildung, Jugend und Sport Berlin, 2006, Anfangsoptik „Sehen und gesehen werden“, Modul P4 7/8

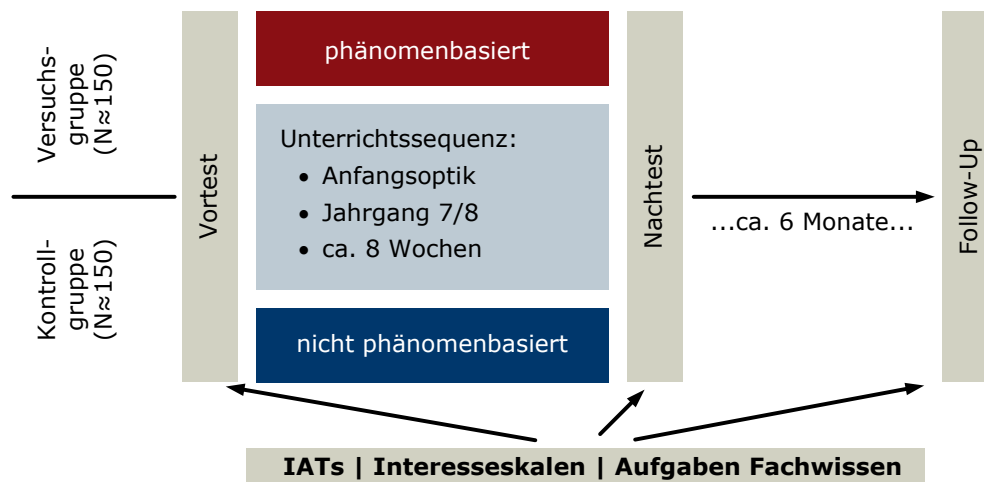


Abb. 7.1 Geplantes Forschungsdesign. Der Follow-Up-Test wurde im Verlauf der Studie letztendlich nicht durchgeführt.

Faktoren müssten nachträglich unter schulischen Bedingungen nachgewiesen werden. Außerdem liegt das Forschungsinteresse dieser Arbeit zunächst nicht darin, die genauen Faktoren PbPUs zu identifizieren, die einen möglichen Einfluss verursachen. Eine solche Konfundierung aufzulösen, wäre das Ziel von Nachfolgeprojekten. Aus diesem Grund fällt die Wahl der Untersuchungsumgebung auf das *Feld*.

Das Untersuchungsfeld *Schule* lässt wenig Möglichkeiten zu, die Untersuchung experimentell, d. h. mit zufällig ausgewählten Stichproben durchzuführen. Bei einer mehrwöchigen Unterrichtssequenz, die von speziell geschulten Lehrern durchgeführt werden muss, können die Probanden nicht zufällig zur Versuchs- bzw. Vergleichsgruppe zugeordnet werden, zumal dieselbe Lehrkraft aus Objektivitätsgründen nicht beide Gruppen zugleich unterrichten soll. Somit ist man auf eine klassenweise Zuordnung angewiesen (*quasiexperimentell*). Dem dadurch entstandenen Mangel an interner Validität, also der möglichen Ursachenzuschreibung auf die unabhängige Variable (Versuchs- vs. Kontrollgruppe), wird durch Kontrolle von Störvariablen und Pre-/Posttestmessungen begegnet.

Damit ist das Design auf eine quasiexperimentelle Vergleichsstudie im Untersuchungsfeld *Schule* festgelegt, in der die abhängigen Variablen in einem Vor- und Nachtest erhoben werden. Um mögliche Veränderungen besser einschätzen zu können, ist es außerdem sinnvoll, nach einer längeren Zeit einen weiteren Nachtest (Follow-Up) durchzuführen (siehe Abb. 7.1). Im Verlauf der vorliegenden Studie hat sich jedoch herausgestellt, dass der ursprünglich geplante Follow-Up-Test nicht sinnvoll gewesen wäre, weshalb auf ihn letztendlich verzichtet wurde (siehe Abschnitt 10.2).

Für die teilnehmenden Lehrer der Kontrollgruppe muss sichergestellt sein, dass sie nie die phänomenbasierte Anfangsoptik, deren Prinzipien oder deren nähere Vorgehensweise kennengelernt haben. Die Zusammensetzung der Stichprobe wird ausführlich in

Abschnitt 10.1 beschrieben. Die Instrumente für die Erhebung der abhängigen Variablen (Interesse, Image, Selbstkonzept, Fachwissen) werden in Abschnitt 7.2 vorgestellt.

Die bisher formulierten Forschungshypothesen können nun auf das Untersuchungsdesign angepasst formuliert werden:

H 1: Nach einer mehrwöchigen Intervention in Form eines phänomenbasierten Unterrichts zur Anfangsoptik verändert sich das Interesse am Physikunterricht und seinen Inhalten positiv im Vergleich zur Kontrollgruppe.

H 2: Nach einer mehrwöchigen Intervention in Form eines phänomenbasierten Unterrichts zur Anfangsoptik verbessert sich das Image von Physik im Vergleich zur Kontrollgruppe.

H 2.a: Nach einer mehrwöchigen Intervention in Form eines phänomenbasierten Unterrichts zur Anfangsoptik sind die Assoziation von *Physik* mit *schwierig* mehr geschwächt als in der Kontrollgruppe.

H 2.b: Nach einer mehrwöchigen Intervention in Form eines phänomenbasierten Unterrichts zur Anfangsoptik sind die Assoziation von *Physik* mit *männlich* mehr geschwächt als in der Kontrollgruppe.

H 3: Eine mehrwöchige Intervention in Form eines phänomenbasierten Unterrichts zur Anfangsoptik wirkt sich besser auf das physikbezogene Selbstkonzept aus, als der Unterricht in der Kontrollgruppe. Die Assoziation zwischen dem *Selbst* und *Physik* werden in der Versuchsgruppe mehr gestärkt, als in der Kontrollgruppe.

H 4: Nach einer mehrwöchigen Intervention in Form eines phänomenbasierten Unterrichts zur Anfangsoptik ist der Verlauf des Lernerfolgs (Fachwissen) in Kontroll- und Versuchsgruppe identisch. In beiden Gruppen ist das Fachwissen nach der Intervention höher als vorher.

Die Form der Daten, die man mit dem hier vorgestellten Design erhebt, besitzt eine hierarchische Struktur. Wenn man allgemein davon ausgeht, dass man an verschiedenen Schulen jeweils Klassen mit verschiedenen Lehrern der Versuchs- und Kontrollgruppe zuordnet, gibt es die Subjektebene, die Klassenebene, die Lehrerebene, die Gruppenebene und die Schulebene. Die Analysemethode, die am adäquatesten mit Daten in dieser Form umgeht und erlaubt, auch Interaktionen zwischen den Ebenen aufzudecken, ist die Mehrebenenanalyse (vgl. Multilevel Linear Models Field, 2009, S. 725ff). Die Vorteile liegen vor allem darin, dass sie weniger Vorannahmen trifft (z. B. Homogenität der Regressionsanstiege und Unabhängigkeit der Daten verschiedener Subjekte), als die sonst üblichen Verfahren (ANOVA, ANCOVA). Außerdem stellen fehlende Daten und somit unausgeglichene Designs für sie kein Problem dar. Ein Nachteil ist die benötigte Stichprobengröße. Deren Berechnung ist sehr komplex, aber laut „Faustformel“ sind auf

höchster Ebene mindestens zehn Einheiten nötig (Kuhn, 2014). Bei einer Beschränkung auf drei Ebenen würde dies im Rahmen der vorliegenden Studie zehn verschiedene Lehrer bedeuten. Hinzu kommt die Tatsache, dass die Ergebnisse der Mehrebenenanalysen selten im Widerspruch zu denen der alternativen Verfahren liegen, da letztendlich allen das Modell der linearen Regression zugrunde liegt. Damit steht der hohe Aufwand in keinem Verhältnis zu den inhaltlichen Ergebnissen (Schnell, Hill & Esser, 1999).

Die am häufigsten angewendeten Verfahren bei Prä-Post-Kontrollgruppendesigns sind varianzanalytische Methoden (ANOVAs)². Passend zum hier vorgestellten Design kommen dafür zunächst eine zweifaktorielle ANOVA mit Messwiederholung auf einem Faktor oder eine einfaktorielle ANOVA der Gain-Scores infrage. Die Ergebnisse beider Verfahren sind absolut identisch (Dimitrov & Rumrill, 2003; Dugard & Todman, 1995), weshalb hier nur Ersteres weiter betrachtet wird. Der erste Faktor wäre der zweistufige Zwischensubjekt-Faktor: Versuchs- oder Kontrollgruppe. Der zweite (Messwiederholungs-)Faktor wäre die (Mess-)Zeit und hätte je nach Anzahl der Tests zwei oder drei Stufen. Die zweifaktorielle ANOVA (Messwiederholung auf einem Faktor) wird üblicherweise für die Auswertung solcher Split-Plot-Designs empfohlen (z. B. Bortz & Döring, 2002; Eid, Gollwitzer & Schmitt, 2011). Ein signifikanter Haupteffekt auf dem Zwischensubjektfaktor würde bedeuten, dass sich unabhängig vom Messzeitpunkt beide Gruppen in der entsprechenden Variable unterscheiden. Dementsprechend würde ein signifikanter Haupteffekt auf dem Messwiederholungsfaktor darauf hindeuten, dass unabhängig von der Gruppe die Variable von der Messzeit abhängig ist. Für die Forschungsfrage wären beide Effekte nur wenig relevant, im Gegensatz zu der Interaktion zwischen beiden Faktoren. Ein signifikanter Interaktionseffekt würde bedeuten, dass die Veränderung der Variablen über die Zeit für beide Gruppen unterschiedlich ist, was bezüglich der Forschungsfragen und Hypothesen sinnvoll interpretiert werden kann.

An verschiedenen Stellen wird darauf aufmerksam gemacht, dass die ANOVA mit Messwiederholung auf einem Faktor für das auch hier verwendete Design nicht das adäquate Analyseverfahren ist (Dimitrov & Rumrill, 2003; Dugard & Todman, 1995; Rausch, Maxwell & Kelley, 2003). Stattdessen wird eine Kovarianzanalyse (ANCOVA) empfohlen, in welcher der Vortest als Kovariate eingeht und der Nachtest³ zwischen den entsprechenden Gruppen verglichen wird. Die ANCOVA hat eine höhere Teststärke, weil sie die Varianz im Vortest als systematischen Fehler für die Auswertung im Nachtest berücksichtigt und somit dort mehr Varianz aufklären kann. Dieser Vorteil geht aber mit strengeren Vorannahmen einher. Zu den Vorannahmen der ANOVA kommt hinzu, dass die Kovariate, also das Vortestergebnis, von der Gruppenzugehörigkeit (bezüglich derer verglichen werden soll) unabhängig ist. Im hier vorliegenden Fall bedeutet dies, dass die Vortestergebnisse sich in Versuchs- und Kontrollgruppe nicht signifikant voneinander unterscheiden dürfen (Field, 2009). Diese Vorannahme ist eine häufig missverstandene oder sogar nicht beachtete Bedingung, wenn z. B. suggeriert wird, dass eine

²ANalysis Of VAriance

³nicht der Zuwachs zwischen Vor- und Nachtest (Gain-Score)

Störvariable in Form einer Kovariaten „kontrolliert“ wird (eine detaillierte Analyse, unter welchen Umständen dies problematisch ist, liefern Miller & Chapman, 2001). Eine weitere zusätzliche Vorannahme ist die Homogenität der Regressionsanstiege in allen zu vergleichenden Gruppen. Beide Vorbedingungen sind in der Regel gegeben, wenn es sich um eine experimentelle Studie handelt, da dort die Gruppenzugehörigkeit per Zufall entschieden wird. In einer quasiexperimentellen Untersuchung müssen sie jedoch unbedingt überprüft werden.

Die ANOVA (mit Messwiederholung) ist als Analysemethode in der vorliegenden Arbeit die erste Wahl. Aufgrund ihrer geringeren Teststärke führt sie zu eher konservativen Entscheidungen bezüglich der Nullhypothese. Für die Variablen, bei denen durch Überprüfung die Gültigkeit der Vorannahmen gewährleistet ist, wird auch eine ANCOVA durchgeführt. Anhand ihrer Ergebnisse können die der ANOVA entweder bekräftigt oder relativiert werden.

7.2 Wahl der Instrumente

Die Forschungsfragen 1-4 bestimmen zunächst die vier abhängigen Variablen *Interesse*, *Image*, *Selbstkonzept* und *Fachwissen*, die noch operationalisiert werden müssen. Das *Image* beinhaltet im Rahmen dieser Arbeit die Dimensionen *Schwierigkeit* und *Geschlechtskonnotation*. Im Folgenden werden die Testinstrumente vorgestellt, die hier zum Einsatz kommen sollen. Aus ihnen ergeben sich noch weitere Subvariablen, die ebenfalls aufgeführt werden.

Impliziter Assoziationen Test

Die Operationalisierung des Images von Physik stellen im Rahmen der BIT Assoziationen von *Physik* zu *schwierig/leicht* und *männlich/weiblich* dar. Die Operationalisierung des Selbstkonzepts geschieht durch Messung der Assoziation zum *Selbst*. Für die Erhebung dieser (impliziten) Assoziationen kommen hier drei IATs zum Einsatz.

Das Verfahren selbst und die Begründung für die Wahl, die sich aus der Testgüte ergibt, ist in Abschnitt 4.1 dargestellt. Alle IATs haben das Unterrichtsfach als Zieldimension mit den Kategorien *Physik* und *Englisch*. Als gegensätzliche Kategorie zu *Physik* wird den Empfehlungen von Lane et al. (2007) entsprechend dasselbe Unterrichtsfach gewählt, wie in den Image-IATs von Kessels et al. (2006). Die Attributdimensionen *Schwierigkeit*, *Geschlechtskonnotation* und *Selbstkonzept* haben die Kategorien *schwierig/leicht*, *Junge/Mädchen* und *Andere/Ich* (siehe Tabelle A.1 im Anhang). Die zugehörigen Stimuli sind alle Wörter (Tabelle A.1). Sie wurden bei der Zieldimension durch eine offene Befragung in der Zielpopulation generiert.⁴ Die Stimuli für die Image-

⁴ „Was verbindet ihr gedanklich mit *Physik/Englisch*?“

dimensionen sind teilweise aus der Studie von Kessels et al. (2006) und teilweise selbst bestimmt, die für das Selbstkonzept entstammen von Schnabel et al. (2007).

Da für die Erhebung aller Variablen drei IATs hintereinander durchgeführt werden müssen, ist der Ablauf der Blöcke (siehe Abb. 4.2) leicht verändert. In allen Tests bleibt die Zieldimension konstant. Deshalb braucht sie auch nur einmal geübt zu werden, sodass Block 1 in den zwei folgenden IATs entfällt. In diesem Fall ist es auch sinnvoll, dass für die inkompatible Paarung, abweichend zu Abb. 4.2, die Attributdimension die Seite wechselt, da nach diesem Wechsel ein erneuter Übungsblock nötig ist.

Die Übungsblöcke bestehen jeweils aus 20 Trials, die Testblöcke aus 40 Trials. Die Stimuli erscheinen randomisiert, jedoch wiederholen sie sich erst, nachdem jeder einmal erschienen ist. Bei fünf verschiedenen Wörtern stellt dies sicher, dass pro Block jedes gleich häufig erscheint, was für die Bestimmung der Reliabilität (siehe 10.4) wichtig ist. Das Assoziationsmaß wird durch den *D*-Wert-Algorithmus berechnet und für den Umgang mit falsch beantworteten Trials wird ein Built-In-Error-Penalty gewählt (siehe Abschnitt 4.1.2). Für die Durchführung der IATs wird Inquisit 3.0.4.0 [Computer Software] verwendet. Der vollständige Algorithmus und das SPSS-Skript der *D*-Wert-Berechnung befinden sich in Anhang A.1 und A.1.

Single-Target IAT

Neben den IATs kommen in der Pilotstudie zur Operationalisierung der oben genannten Konstrukte auch Single-Target IATs zum Einsatz. Bei ihnen entfällt die zweite Zielkategorie *Englisch*. Alle weiteren Kategoriebezeichnungen und Stimuli entsprechen denen der IATs (Tabelle A.1). In der ersten Pilotierung bestanden die Übungsblöcke aus 20 Trials und die Testblöcke aus 35 Trials. In der zweiten Pilotierung wurde die Trialanzahl variiert. Für das Seitenverhältnis der Tastendrucke wird der von Bluemke und Frieze (2008) vorgeschlagene Kompromiss 2 : 2 : 3 gewählt (siehe Abschnitt 4.2). Das Assoziationsmaß wird ebenfalls durch den *D*-Wert-Algorithmus berechnet. Der vollständige Algorithmus des ST-IATs befindet sich in Anhang A.1.

Explizite Assoziationen

Parallel zu den impliziten Tests werden alle Assoziationen auch mit expliziten Skalen erhoben. Sie dienen der zusätzlichen Validierung und haben die Form von semantischen Differentialen. Zu jedem Itemstamm gibt es eine quasi-stufenfreie Skala zwischen zwei Antwortpolen. Bei der Erhebung am Computer zusammen mit den IATs kann für jede Antwort ein Schieberegler zwischen den Polen positioniert werden. Die Codierung der Position erfolgt mit Zahlen zwischen 0 und 10, wobei mit 5 die neutrale Position in der Mitte gekennzeichnet ist.

Zu jeder Attributdimension im IAT oder ST-IAT gibt es zwei semantische Differentiale:

Schwierigkeit Das Fach *Physik* finde ich eher leicht... ...schwierig

	Das Fach <i>Physik</i> ist eher	schwer verständlich...	...leicht verständlich
Geschlechts-	<i>Physik</i> verbinde ich eher mit	Jungen...	...Mädchen
konnotation	Das Fach <i>Physik</i> ist eher ein	„Mädchen-Fach“...	..., „Jungen-Fach“
Selbstkonzept	Das Fach <i>Physik</i> passt	zu mir...	...nicht zu mir
	Physikunterricht ist eher etwas für	andere...	...mich

Interessefragebogen

Für die Erhebung des Interesses wird ein Fragebogen aus Skalen zusammengestellt, die bei der PISA-Studie von 2006 eingesetzt wurden (OECD, 2007a; Frey et al., 2009). Die Wahl fällt auf diese erprobten Skalen, weil sie eine hohe Testgüte aufweisen. Außerdem können die hier erhobenen Daten mit denen der PISA-Studie verglichen werden. Bezüglich der Reliabilität und der Validität braucht dieses Instrument demnach nicht weiter diskutiert werden. Die hier verwendeten Skalen sind in Tabelle 7.1 aufgeführt. Die bisher nur allgemein als *Interesse* bezeichnete abhängige Variable gliedert sich also bei einer Auswertung auf Skalenebene in sechs bzw. bei einer Auswertung auf Subskalenebene in zehn Subvariablen auf. Die Skalen *Unterstützende Lehr-/Lern-Bedingungen* und *Qualität der Lernmotivation* werden im Folgenden mit den Kürzeln *wsewkuwau* und *mo123* bezeichnet. Der vollständige Interessefragebogen befindet sich im Anhang A.2.

Test zum Fachwissen

Wie in Abschnitt 6.1 dargelegt wurde, soll der Fachwissenstest hauptsächlich kontrollieren, ob für die Schülerinnen und Schüler in der Versuchsgruppe durch den Unterricht der phänomenbasierten Anfangsoptik ein Nachteil in den erworbenen fachlichen Kompetenzen entsteht. Daher lautet die Anforderung an den Test, dass er möglichst inhaltsvalide das Fachwissen prüft, welches Schülerinnen und Schüler nach Absolvieren der Anfangsoptik erworben haben sollten, unabhängig davon, auf welche Art sie unterrichtet wurden.

Aus diesem Grund besteht der hier zum Einsatz kommende Test aus Aufgaben aus dem Itempool zur Evaluation der Standards in den Naturwissenschaften (ESNaS), die mithilfe der fachdidaktischen Leitungen unter Obhut des Instituts zur Qualitätsentwicklung im Bildungswesen (IQB) entwickelt wurden (siehe Walpuski et al., 2010). Die Aufgaben wurden sowohl unter fachdidaktischer als auch psychometrischer Perspektive sehr aufwändig entwickelt. Die fachdidaktische Betreuung der Entwicklung stellt die geforderte Validität sicher. Denn es kann davon ausgegangen werden, dass die Inhalte der Aufgaben als gemeinsamer Nenner dessen gelten, was am Ende der Sekundarstufe I aus fachdidaktischer Sicht in Physik an Fachwissen erworben werden sollte. Dies schafft die Voraussetzung für einen fairen Vergleich zwischen Versuchs- und Kontrollgruppe. Auch die anderen psychometrischen Eigenschaften der Aufgaben sind sehr gut. Sie sind pilotiert, rasch-homogen und an einer großen Eichstichprobe normiert worden.

Von den Aufgaben aus dem Bereich *Optik* wurden elf Items aus den Kompetenzbereichen *Fachwissen* und *Erkenntnisgewinnung* ausgewählt, die die Inhalte der An-

Tabelle 7.1 Übersicht über die eingesetzten Interessesskalen mit Beispielitems.

Skala	Subskala	Kürzel	Beispielitem
Unterstützende Lehr-/Lern-Bedingungen	Wahrgenommene Soziale Eingebundenheit	wse	Die Lehrerin/ der Lehrer schätzt meine Mitarbeit im Unterricht.
	Wahrgenommene Kompetenzunterstützung	wku	Die Lehrerin/ der Lehrer traut mir anspruchsvolle Dinge zu.
	Wahrgenommene Autonomieunterstützung	wau	Die Lehrerin/ der Lehrer ermuntert uns, selbst zu überlegen, wie man am besten vorgeht.
Freude und Interesse an Naturwissenschaften		JOYSCIE	Im Allgemeinen macht es mir Spaß, mich mit naturwissenschaftlichen Themen zu befassen.
Instrumentelle physik-bezogene Motivation		INSTSCIE	Ich lerne Physik, weil ich weiß, dass es für mich nützlich ist.
Interesse über Naturwissenschaften zu lernen		INTSCIE	Wie hoch ist dein Interesse, etwas über Themenbereiche in der Astronomie zu lernen?
Qualität der Lernmotivation	Amotiviertheit/Externale Motiviertheit in Physik	mo1	Im Unterricht in Physik habe ich keine Lust mich zu beteiligen.
	Introjierte/Identifizierte Motiviertheit in Physik	mo2	Im Unterricht in Physik arbeite ich mit, weil ich es immer so mache.
	Intrinsische/Interessierte Motiviertheit in Physik	mo3	Im Unterricht in Physik bin ich mit Freude dabei.
Wahrgenommene Bedeutung der Lerninhalte in Physik		meco	Im Unterricht in Physik wird erkennbar, wofür die Experimente nützlich sind.

fangsoptik betreffen. Dabei wurde darauf geachtet, dass die Schwierigkeiten der Aufgaben nicht zu hoch sind.⁵ Ersteres geschah vor dem Hintergrund, dass zwischen beiden Kompetenzbereichen hier nicht weiter unterschieden werden soll. Das Konstrukt wird deshalb einheitlich mit *Fachwissen* bezeichnet. Bei der durchzuführenden Testung geht es weniger um eine (absolute) Verortung in einem Kompetenzstrukturmodell, als um einen Vergleich zweier Lerngruppen. Dieser soll möglichst viele Aspekte eines zur Optik spezifischen Lernerfolg berücksichtigen. Die Aufgaben mit zu hohen Schwierigkeiten wurden deshalb nicht ausgewählt, weil die Zielpopulation im ESNaS-Projekt eine andere als die in der vorliegenden Studie ist. Der dem ESNaS-Projekt zugrunde liegende angestrebte Ländervergleich zielt auf den mittleren Schulabschluss, also das Ende der Sekundarstufe I (Klassenstufe 10). Die hier interessierende Population befindet sich größtenteils in Klassenstufe 7.

Die Aufgaben umfassen die Inhalte *Modell eines Lichtstrahls*, *Gegenstände sehen*, *Reflexion am ebenen Spiegel*, *Schatten* und *leuchtende und beleuchtete Körper*. Auf eine vollständige Darstellung der Items muss an dieser Stelle verzichtet werden, weil sie aufgrund der zukünftig durchzuführenden Ländervergleichsstudien nicht öffentlich zugänglich sein dürfen.

7.3 Statistische Hypothesen

Nachdem das Design, die Analyseverfahren und die Testinstrumente festgelegt sind, können die darauf angepassten statistischen Hypothesen formuliert werden:

SH 1: Bei den Mittelwerten der Interessesskalen besteht eine signifikante Interaktion zwischen dem Faktor *Testzeit* und dem Faktor *Gruppe*. Der Verlauf der Skalenmittelwerte über die Zeit hat in der Versuchsgruppe einen größeren Anstieg als in der Kontrollgruppe.

SH 2: Bei den Assoziationsmaßen zum Image von Physik besteht eine signifikante Interaktion $Testzeit \times Gruppe$.

SH 2.a: Vor der Intervention bestehen unabhängig von der Gruppe positive *D*-Werte in den Dimensionen *Schwierigkeit* und *Geschlechtskonnotation*, d. h. eine Assoziation von *Physik* mit *schwierig* und *männlich*.

SH 2.b: Für den *D*-Wert der Dimension *Schwierigkeit* (D_{Schw}) besteht eine signifikante Interaktion $Testzeit \times Gruppe$. Der Verlauf von D_{Schw} über die Zeit hat in der Versuchsgruppe einen kleineren Anstieg als in der Kontrollgruppe.

SH 2.c: Für den *D*-Wert der Dimension *Geschlechtskonnotation* (D_{Ge}) besteht eine signifikante Interaktion $Testzeit \times Gruppe$. Der Verlauf von D_{Ge} über die Zeit hat in der Versuchsgruppe einen kleineren Anstieg als in der Kontrollgruppe.

⁵Für die Einschätzung lagen die sich aus der Raschskalierung ergebenden Aufgabenparameter der Normierung zur Verfügung.

SH 2.d: Für den Mittelwert der semantischen Differentiale in der Dimension *Schwierigkeit* besteht eine signifikante Interaktion *Testzeit* \times *Gruppe*. Der Verlauf des Mittelwerts über die Zeit hat in der Versuchsgruppe einen kleineren Anstieg als in der Kontrollgruppe.

SH 2.e: Für die Mittelwerte der semantischen Differentiale in der Dimension *Geschlechtskonnotation* besteht eine signifikante Interaktion *Testzeit* \times *Gruppe*. Der Verlauf des Mittelwerts über die Zeit hat in der Versuchsgruppe einen kleineren Anstieg als in der Kontrollgruppe.

SH 3: Bei den Assoziationsmaßen von *Physik* zum *Selbst* besteht eine signifikante Interaktion *Testzeit* \times *Gruppe*.

SH 3.a: Für den *D*-Wert der Dimension *Selbstkonzept* (D_{Se}) besteht eine signifikante Interaktion *Testzeit* \times *Gruppe*. Der Verlauf von D_{Se} über die Zeit hat in der Versuchsgruppe einen kleineren Anstieg als in der Kontrollgruppe.

SH 3.b: Für die Mittelwerte der semantischen Differentiale in der Dimension *Selbstkonzept* besteht eine signifikante Interaktion *Testzeit* \times *Gruppe*. Der Verlauf des Mittelwerts über die Zeit hat in der Versuchsgruppe einen kleineren Anstieg als in der Kontrollgruppe.

SH 4: Der Score des Fachwissenstests entwickelt sich in Versuchs- und Kontrollgruppe gleich.

SH 4.a: Für den Score des Fachwissenstests besteht ein signifikanter Haupteffekt beim Faktor *Testzeit*. Der Verlauf des Scores über die Zeit ist über beide Gruppen hinweg positiv.

SH 4.b: Für den Score des Fachwissenstests besteht keine signifikante Interaktion *Testzeit* \times *Gruppe*. Der Verlauf des Scores über die Zeit hat in der Versuchsgruppe den gleichen Anstieg wie in der Kontrollgruppe.

KAPITEL 8

Treatment

In dem folgenden Kapitel wird das Treatment dargestellt. Dies soll dem Leser ermöglichen, den in den Versuchsgruppen stattgefundenen Unterricht in Inhalt, Zeit und Umfang einschätzen zu können. Das Kapitel ist in der Vergangenheitsform geschrieben, weil es nicht die geplante Unterrichtssequenz vorstellt, sondern den tatsächlich stattgefundenen Verlauf wiedergibt.

Die phänomenbasierte Anfangsoptik wurde in der Hauptstudie von vier verschiedenen Lehrern in fünf Klassen unterrichtet (siehe Tabelle 10.1, S. 130). Aufgrund unterschiedlicher Anzahl an Teilungsstunden und anderen Abweichungen der äußeren Bedingungen war die Planung der einzelnen Unterrichtseinheiten nicht in allen Klassen identisch. Selbst in den beiden Klassen, die vom selben Lehrer an derselben Schule unterrichtet wurden (Lehrer NW, Klassen 4 und 7)¹, gab es Unterschiede, weil mit Beginn der Studie die Reihenfolge geteilte Klasse / gesamte Klasse nicht dieselbe war. Der grobe Ablauf ist jedoch in allen Gruppen vergleichbar. Die Schlüsselstellen des Unterrichts und die Merkmale PbPUs waren in allen Gruppen vertreten. Um dies sicherzustellen, wurde der nicht vom Autor gehaltene Unterricht stichprobenartig hospitiert und die Unterrichtsplanung eingesehen. An dieser Stelle wird beispielhaft für alle Lerngruppen der Unterricht in Klasse 4 dargelegt.

Die Unterrichtssequenz bestand aus insgesamt 14 Unterrichtseinheiten à 45 min. Davon waren sechs Einheiten Teilungsstunden und in acht Stunden wurde die gesamte Klasse gemeinsam unterrichtet. Da schon zu Beginn abzusehen war, dass für das Optikmodul keine 20 Unterrichtseinheiten zur Verfügung stehen werden, musste von der Planung abgewichen werden, die von Weber und Schön (2000) vorgestellt wird. Die Betrachtungen zur Form und zum Lauf des Mondes am Anfang sowie die Totalreflexion

¹Eine genaue Beschreibung der Stichprobe erfolgt in Abschnitt 10.1.

und die Schusterkugel am Ende wurden ausgelassen. Bei der Darstellung der Anfangsoptik in Abb. 2.4 (S. 33) entspricht dies dem Verzicht auf die letzte Kachel *Totalreflexion*. Die erste Kachel *Das Licht und die Dinge* wird in verkürzter Form berücksichtigt.

1. Stunde: Das Wesen des Lichts und der weiße Schatten

Die erste Stunde diente der Einführung in die Optik als *Lehre vom Licht*. Jedoch musste zunächst demonstriert werden, warum es nicht zweckmäßig ist, von Beginn an das *Wesen des Lichts* zu thematisieren. Das geschah in zwei Schritten: Zunächst wurden Begriffe gesammelt, die von den Schülerinnen und Schülern mit *Optik* in Verbindung gebracht werden. Bereits hier stellt sich heraus, dass das Genannte nur sehr selten *Licht* war und häufiger etwas mit Brillen, Kameras, Projektoren oder Linsen zu tun hatte. Im zweiten Schritt sollte in kleinen Gruppen überlegt werden, wie man einem Austauschschüler das Wort *Licht* verständlich macht, wenn man die Vokabel dafür nicht kennt. Für gewöhnlich zeigt man in solchen Fällen auf den entsprechenden Gegenstand und die Schülerinnen und Schüler sollten demonstrieren, wie sie vorgehen würden. Sie hatten dafür verschiedene Lichtquellen, wie z. B. Taschenlampen oder Teelichter zur Verfügung. Den Vorschlägen der Gruppen konnte vom Lehrer immer begegnet werden, indem er anmerkte, dass stets auf Leuchten oder beleuchtete Flächen gezeigt wurde. Die einzige Möglichkeit dies zu umgehen bestand in der Umschreibung, dass Licht das sei, was aus der Lampe „heraus kommt“. Zeigen konnte es niemand.

Im Anschluss daran folgte ein Demonstrationsversuch im abgedunkelten Raum mit einer Taschenlampe. Von den Schülerinnen und Schülern wurde von der Seite nur die Leuchte und der beleuchtete Gegenstand gesehen, dazwischen nichts. Außerdem wurde Kreidestaub zwischen Lampe und Wand gebracht. Dies erweckte zunächst den Eindruck, man könnte das Lichtbündel nun sehen. Bei genauer Betrachtung bemerkten jedoch alle, dass dies auch nur der Kreidestaub war, den man erleuchtet sah. Es folgte der erste Merksatz für die Optik:

Merksatz 1: Licht ist unsichtbar: Wir sehen beleuchtete Gegenstände oder Leuchten, aber nie das Licht auf seinem Weg zwischen zwei Punkten. (Mikelskis et al., 2006, S. 89)

Aus dieser unmittelbar von jedem nachvollziehbaren Tatsache heraus wurde vom Lehrer festgehalten, dass es für die kommenden Stunden weniger um das Licht und sein Wesen gehen sollte. Schließlich sei es schwierig zu untersuchen, da es unsichtbar ist. Stattdessen sollte vielmehr jenes in das Zentrum der Betrachtungen geraten, was sichtbar ist: die gesehenen beleuchteten Gegenstände (Bilder) und die Leuchten selbst.

Als erstes Beispiel für dieses zukünftige Vorgehen wurde das Experiment zum *weißen Schatten* mit dem Würfelkantengerüst demonstriert (siehe Schilderung des Experiments ab S. 38). Schließlich wurden der Aufbau des Experiments und die Beobachtung an der Tafel festgehalten.

2. Stunde: Das *Prinzip Ameise*

Der seltsame und bisher unerklärliche *weiße Schatten* wurde zu Beginn der Stunde wieder aufgegriffen und zu Erinnerung nochmals mit dem Würfelkantengerüst demonstriert. Für die weiteren Untersuchungen wurde der Aufbau jedoch nun vereinfacht und damit leichter beschreibbar. Anstelle des Gerüsts als Schattengeber wurde ein breites, senkrecht montiertes Tafellineal als Schattengeber benutzt. Der Rest des Experiments blieb gleich. Auch in der neuen Konstellation bildete sich ein weißer Schatten des Lineals ab, der durch Bewegung der Finger vor der Stableuchte schnell erkannt wurde.

Der erste Schritt sich dem Phänomen zu nähern, geschah durch eine stufenlose Variation der Stableuchte, indem ihre Länge verändert wurde. Rechts und links wurden Pappröhren über die Leuchte dicht zusammengeschoben. Ohne die Finger als zusätzliche Blende wurde zunächst der scharfkantige *schwarze Schatten* des Lineals projiziert. Die Lichtquelle konnte nun nach und nach von einer Punktleuchte zu einer ausgedehnten Leuchte transformiert werden, unter gleichzeitiger Beobachtung der Entwicklung des Schattenbildes. Die dabei festgehaltene Erkenntnis lautete, dass die Form und Helligkeit des Schattens von der Form der Leuchte abhängt. Dabei wurden die Abstände so gewählt, dass in der Mitte des Schattenbereichs am Ende der Transformation der Leuchte ein dunkler Kernschatten zu sehen war.

Mit dem Ziel, herauszufinden, in welchem Maße diese Abhängigkeiten zwischen Leuchte, Schatten und Schattengeber bestehen, wurde eine Untersuchungsmethode eingeführt, die bei zukünftigen Betrachtungen immer wieder angewendet werden sollte: Das *Prinzip Ameise* (siehe S. 33 ff). Jede Schülerin und jeder Schüler sollte den Gang einer Ameise an der Wand einmal nachvollziehen und die Ansicht der Lampe vom Ort an der Wand mit dessen Helligkeit vergleichen. Für diese Unterrichtsphase standen mehrere Aufbauten zur Verfügung und die Lernenden konnten frei und selbstständig experimentieren und die Abstände oder die Form der Lampe variieren.

Im Anschluss an die Experimentierphase wurde in Stillarbeit ein Arbeitsblatt zum PA (siehe B.1 im Anhang) bearbeitet und am Ende der Unterrichtsstunde verglichen.

3. Stunde: Schülerexperiment zum *Prinzip Ameise*

Der nächste Schritt zu Klärung des weißen Schattens geschah in Form von angeleiteten Schülerexperimenten (siehe Aufgabenblatt B.2 im Anhang). Mithilfe von zwei Teelichtern, einer Streichholzschachtel und einem DIN A4-Blatt Papier sollten drei verschiedene Schattenprojektionen realisiert werden und die Sichtposition der Ameise eingenommen werden. Da die Aufbauten nur zwei Lichtquellen beinhalteten, gab es auch nur diskrete Schattenbereiche, deren Helligkeiten mit den Ansichten der Ameise (Anzahl der gesehenen Kerzen) tabellarisch verglichen werden konnte.

4. Stunde: Photometrischer Grundsatz und Auflösung des weißen Schattens

Nachdem die Beobachtungen und Ergebnisse des Schülerexperiments wiederholt und als Demonstrationsexperiment nachvollzogen wurden, konnte ein Tafelbild zu den Schattenbereichen (Teilschatten und Kernschatten) mit den entsprechenden Sichtbeziehungen entwickelt werden. Dem schließt sich der photometrische Grundsatz „Hell ist es, von wo aus Helles zu sehen ist“ (Grebe-Ellis, 2006b, S. 2) in Form dreier Merksätze an:

- Merksatz 2:**
- a) Eine Fläche erscheint uns umso dunkler, je weniger Leuchten oder helle Flächen von dort aus gesehen werden können.
 - b) Eine Fläche erscheint umso heller, je mehr Leuchten oder helle Flächen von dort aus gesehen werden können.
 - c) Es ist nur dort ganz dunkel, von wo aus weder eine Leuchte noch irgendein beleuchteter Gegenstand zu sehen ist. (Mikelskis et al., 2006, S. 87)

Die eingezeichneten Schattengrenzen wurden als ausgezeichnete *Lichtwege* eingeführt. Dies machte sie ab sofort zu einem Werkzeug, um Lichtwege sichtbar zu machen.

Die Merksätze wurden nun auf den nicht mehr diskreten Fall der Stabileuchte angewendet. Mit dem Aufbau vom Beginn der Stunde konnte demonstriert werden, dass der Schatten im Übergangsbereich deshalb im Verlauf zum Rand heller wird, weil von ihm aus gesehen der Schattengeber immer weniger der Leuchte verdeckt. Zusätzlich wurde eine Flash-Simulation gezeigt, die den Vergleich der abgelösten und eingebundenen Perspektive für alle Schülerinnen und Schüler ermöglicht (siehe Abb. 8.1).

Mithilfe der Merksätze konnte abschließend auch das Phänomen des weißen Schattens verstanden werden. Nach Merksatz 2 muss vom Ort des weißen (hellen) Schattens mehr von der Stabileuchte sichtbar sein, als von den benachbarten Bereichen. Mit dem gleichen Aufbau, der bisher verwendet wurde, konnte nun auch demonstriert werden, dass, vom hellen Schatten aus geschaut, der Schattengeber die beiden Finger vor der Leuchte verdeckt und somit mehr von ihr zu sehen ist, als in den Bereichen, in denen Finger *und* Schattengeber sie verdecken. Diese Erkenntnis wurde ebenfalls mit einem Tafelbild festgehalten.

In dieser Stunde wurde am Ende eine Hausaufgabe erteilt. Die Schülerinnen und Schüler sollten zur übernächsten Stunde (nach einer Woche) eine selbstgebaute Lochkamera mitbringen. Für die Bauanleitung hatten sie neben der Internetrecherche Mikelskis et al. (2006, S. 92) zur Verfügung.

5. Stunde: Schattenformen verschiedener Leuchten und Schattengeber

Da zu Beginn der fünften Stunde das Rätsel des weißen Schattens mithilfe der bereits erarbeiteten Gesetzmäßigkeiten gelöst war, sollte nun der bereits beobachtete Zusammen-

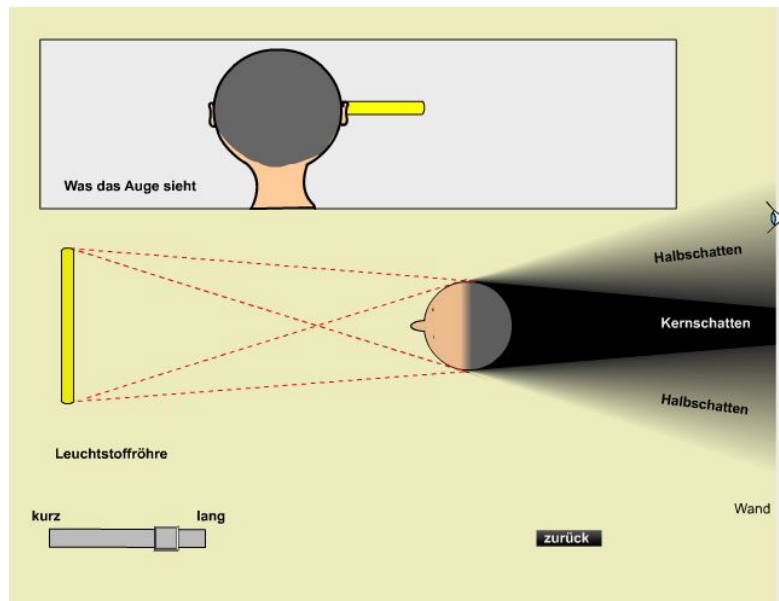


Abb. 8.1 Flash-Simulation zur abgelenkten und eingebundenen Perspektive bei der Schattenprojektion einer Stablampe. Die Länge der Leuchte sowie der Ort auf dem Schirm kann vom Benutzer stufenlos variiert werden. (Cornelsen Schulverlage GmbH, 2014)

hang zwischen Schattenform und Form der Lampe bzw. des Schattengebers in den Fokus gelangen. Dazu wurden in einem Demonstrationsexperiment systematisch die Leuchten und Schattengeber variiert und die Beobachtungen übersichtlich dargestellt. Die Schülerinnen und Schüler hatten dafür ein Arbeitsblatt (B.3 im Anhang) zur Hilfe.

Zunächst wurden mit einer Punktleuchte drei verschiedene Schattengeber beleuchtet. Ein aus Stativstangen gebautes Kreuz, ein an einem Draht befestigtes gleichseitiges Dreieck, Spitze nach unten, mit einer Kantenlänge von ca. 1 cm und eine schwarze Pappe im DIN-A2-Format, mit einem Loch der gleichen Form in der Mitte, wie das verwendete Dreieck. An der Projektionsfläche waren nacheinander scharfkantige Schatten zu beobachten, die alle die Form des jeweiligen Schattengebers hatten: ein dunkles Kreuz, ein dunkles Dreieck und ein helles Dreieck, beide mit der Spitze nach unten.

Anschließend wurde eine waagrecht ausgerichtete Stablampe mit denselben Schattengebern verwendet. Beim Stativkreuz war als Projektion zunächst nur ein scharfkantiger waagerechter Strich zu beobachten, der an den Rändern rechts und links kontinuierlich heller wurde. Der senkrechte Strich war zunächst gar nicht zu erkennen. Eine leichte Variation der Abstände führte schließlich dazu, dass auch der Schatten der senkrechten Stange sichtbar projiziert wurde, jedoch unscharf, breitgezogen und hell. Diese Form des Schattens war bereits aus der 2. Stunde bekannt (Tafellineal als Schattengeber). Die Schatten bei beiden Dreiecken hatten die Form eines begrenzten waagerechten Strichs, der erste dunkel, der zweite hell. Die Form der Lampe war hier deutlich zu erkennen.

Die zweite Variation der Leuchte erfolgte, indem sie nun senkrecht montiert wurde. Die Qualitäten der Schattenprojektionen bei allen Schattengebern entsprachen jenen der vorherigen Situation. Die Schatten waren lediglich um 90° gedreht, genauso wie die Leuchte. Damit war der letzte Zweifel ausgeräumt, dass es sich bei der Schattenform um die Form der Lampe handelt.

Den Abschluss der Versuchsreihe bildete eine Ringleuchte, die jedoch nicht für die Schülerinnen und Schüler sichtbar war. Sie konnten lediglich nacheinander die Schattenbilder des Kreuzes sowie der beiden Dreiecke sehen. Anhand der Bilder sollten sie die Form der Leuchte vorhersagen. Der Schatten des Stativkreuzes war sehr komplex und er allein hätte die Vorhersage sehr schwierig gemacht. Beim genauen Hinsehen konnte man allerdings bereits ringförmige Strukturen an den Rändern und im Zentrum des Kreuzes erkennen. Die Schatten der beiden Dreiecke hatten jedoch eindeutig die Gestalt eines dunklen bzw. hellen Ringes. Die Vorhersage der Ringleuchte war nun für die meisten leicht möglich. Es herrschte lediglich Unsicherheit, weil eine derartige Lampenform vielen nicht bekannt war.

Am Ende der Stunde wurde bereits eine Brücke zur Lochkamera gebildet. Zwei Stableuchten wurden L-förmig montiert und als Schattengeber diente die große Pappe mit dem dreieckigen Loch in der Mitte. Wie zu erwarten, hatte der Schatten die Form des L's, jedoch kopfstehend und seitenverkehrt. Mit dieser Beobachtung wurden die Schülerinnen und Schüler entlassen.

6. Stunde: Die Lochkamera

Das Ziel der sechsten Stunde war die Hinführung zur Lochkamera, deren Bilder entsprechend der letzten Stunde als Schattenbilder identifiziert werden sollten. Die selbstgebaute Exemplare der Schülerinnen und Schüler (Hausaufgabe) wurden an diesem Tag mitgebracht.

Die L-förmigen Stableuchten waren wieder aufgebaut. An den Enden beider Schenkel wurden zur besseren Identifizierung eine blaue und eine rote Farbfolie geklebt. Anstelle der Pappe als Schattengeber wurde ein großer undurchsichtiger Vorhang mit einem Loch verwendet. Der Aufbau stand quer zum Auditorium. Das Bild des L's war wie in der letzten Stunde kopfstehend und seitenverkehrt. Dies konnte nun anhand der farbigen Folien auch überzeugender demonstriert werden. Außerdem war jetzt sichtbar, dass die Schattenbilder auch farbig wie die Leuchte sind. Zusätzlich zur Demonstration vorne konnte jeder mit der eigenen Lochkamera das L betrachten und für sich zum selben Ergebnis gelangen.

Nachdem die Beobachtung mit einem Tafelbild festgehalten war, sollte geklärt werden, warum das Bild kopfstehend, seitenverkehrt und farbig ist. Dies geschah durch die Anwendung von *Prinzip Ameise*. Am großen Aufbau konnte vom Ort des Bildes durch das Loch im Vorhang geschaut werden. Vom blauen Ende des senkrechten Schenkels, das sich ganz unten auf der Wand befand, sah man durch das Loch das blau beklebte

Ende der Stablampe, welches ganz oben war. Analog dazu war die Beobachtung des waagerechten Schenkels mit der rechts/links-Vertauschung des roten Endes.

Der letzte Schritt geschah durch den Vergleich der bisherigen Beobachtungen mit denen der eigenen Lochkameras beim Blick aus dem Fenster. An dieser Stelle wurde daran erinnert, dass im Merksatz 2 von Leuchten oder beleuchteten Flächen die Rede ist, weshalb in der Lochkamera auch Häuser oder Bäume sichtbar sind.

Anmerkung: Im Gegensatz zu der hier skizzierten Stunde wird bei der phänomenbasierten Anfangsoptik für gewöhnlich der Unterricht damit begonnen, dass der Klassenraum komplett abgedunkelt und durch ein kleines Loch in der Verdunkelung zur großen begehbaren Lochkamera wird. Damit sind die Beobachtungen noch eindrucksvoller. In den Klassen 1 bis 3 wurde dies auch so durchgeführt. In den Klassen 4 und 7 war es jedoch aufgrund der äußeren Bedingungen nicht möglich.

7. Stunde: Das Doppelschattenexperiment

Das Doppelschattenexperiment bildet die Brücke zwischen den Thematiken *Schatten* und *Spiegel*. Zur Vorbereitung der Stunde wurde der Klassenraum komplett abgedunkelt. Da als Verdunkelung nur lichtdichte Vorhänge zur Verfügung standen, wurden die Fenster zusätzlich mit schwarzem Filzstoff beklebt. Ebenso wurden die Schlitzte unter den Türen und die leuchtenden Schalter der Stromversorgung abgeklebt. Die Sitzordnung wurde zu einem großen Kreis verändert. Einige im Hintergrund befindliche Gegenstände, die stark reflektieren, wurden ebenfalls mit schwarzem Stoff verdeckt. Außerdem wurde eine große Leiter in den Raum gebracht.

In der Mitte des Kreises befand sich ein kleines Podest, welches mit schwarzem Tuch bedeckt war. Auf dem Tuch wurde ein sorgfältig polierter Spiegel gelegt. Über dem Spiegel wurde eine weiße Pappe in der Größe des Podests positioniert. In der Mitte des Kartons war eine kreisrunde Öffnung mit ca. 30 cm Durchmesser. Darauf lag ein weißer Deckel mit ca. 32 cm Durchmesser, ebenfalls aus Pappe. Die Blende und der Deckel waren von unten matt schwarz lackiert.

Unter der Decke des Raumes wurde eine Experimentierleuchte (punktförmig) positioniert, die auf das Podest gerichtet war. Mithilfe einer Irisblende wurde das Lichtbündel so schmal eingestellt, dass nur die weiße Pappe mit Deckel beleuchtet war.

Zu Beginn der Stunde betraten die Schülerinnen und Schüler bei voller Beleuchtung den Raum. Sie sahen bereits das gesamte Arrangement, jedoch nicht, was sich unter der Pappe befand. Nachdem alle auf ihrem Platz waren, wurde das Licht komplett ausgeschaltet. Nach ein paar Minuten, als die Hell-Dunkel-Adaption abgeschlossen war, konnten wieder einige wenige Strukturen im Raum erkannt werden. Schließlich wurde die Experimentierleuchte eingeschaltet und der Deckel minimal zur Seite geschoben. Der darunter befindliche Spiegel zeigte aufgrund der guten Verdunkelung von jedem Sitz nichts als Schwärze, die sogar dunkler war, als das schwarze Tuch auf dem Podest.

Die Schülerinnen und Schüler sollten erraten, was sich unter dem Deckel befand. Wegen der Dunkelheit entstand sogar der Eindruck eines Loches. In wenigen Schritten wurde der Deckel stückweise zur Seite geschoben. Schließlich vermuteten einige einen Spiegel darunter, da sie den sichelförmigen hellen Fleck an der Decke oder im Spiegel sichtbare Gegenstände des Raumes entdeckten.

Als das Geheimnis gelüftet wurde, schlossen sich bereits zwei Fragen bezüglich des Spiegels an, die im künftigen Unterricht weiter thematisiert werden sollten: 1. Welche Farbe hat der Spiegel? 2. Wo ist befindet sich die für den Fleck an der Decke verantwortliche Leuchte?

Im weiteren Verlauf wurden vom Lehrer beide Hände seitlich über dem Podest gehalten. Langsam bewegte er zunächst eine Hand über den Spiegel. Die Schülerinnen und Schüler sollten gleichzeitig die Schatten an der Decke beobachten. Dort erschienen schließlich die Schatten zweier Hände, nachdem der Lehrer seine Bewegung fortführte. Nach kurzer Zeit folgten ein dritter und ein vierter Schatten, was im allgemeine Verwunderung auslöste.

Die Stunde war bis zu diesem Zeitpunkt bewusst wie eine Theater- oder Zaubervorführung inszeniert. Die Eindrücke der absoluten Dunkelheit, die Hervorhebung des Untersuchungsgegenstandes durch einen einzelnen Scheinwerfer, die erstaunlichen Beobachtungen sowie das darbietende Verhalten des Lehrers machten diesen Versuch zu einem Beispiel für die affektive Komponente des phänomenbasierten Unterrichts.

Nun ging es darum zu klären, wie die (doppelten) Schatten zustande kamen. Dafür konnte wieder sachlicher und analytischer vorgegangen werden. Als Schattengeber diente jetzt eine Papp-Hand mit unterschiedlicher Farbe auf beiden Seiten. Auf die Untersuchungsmethode zur Aufklärung kamen die Schülerinnen und Schüler mittlerweile von allein. Um sich an den Ort des Schattens zu begeben, musste eine Schülerin auf die Leiter klettern. Sie folgte der Aufforderung, auf die Leuchte zu zeigen, die sie von dort sehen konnte, wenn sie sich aus dem Schatten heraus begab. Sie zeigte auf die in dem Spiegel für sie sichtbare Experimentierleuchte, die ab jetzt *Spiegelleuchte* genannt wurde. Damit sollte die zweite Frage (siehe oben) teilweise beantwortet werden. Nun sollte sie sich nacheinander in die beiden Schatten begeben und der Klasse nennen, was die Lampe jeweils verdeckt. Die Antwort lautete: 1. eine über dem Spiegel befindliche Hand und 2. eine „im“ Spiegel befindliche Hand (*Spiegelhand*), unterscheidbar durch verschiedene Farben.

Im Prinzip war das Phänomen des Doppelschattens bereits an dieser Stelle verstanden, da die Entstehung mit dem *Prinzip Ameise* nachvollzogen werden konnte. Offen blieb jedoch, wo sich die im Spiegel gesehenen Gegenstände (*Spiegelhand* und *Spiegelleuchte*) befanden.

8. Stunde: Übungen

Da für das weitere Vorgehen Schülerexperimente geplant waren, jedoch eine Stunde mit der gesamten Klasse anstand, wurde eine Stunde mit Übungen durchgeführt. Dies sollte das bisher Gelernte mit dem Ausblick auf die anstehende Lernerfolgskontrolle festigen. Die Schülerinnen und Schüler bearbeiteten selbstständig drei Aufgaben (siehe B.4 im Anhang) zu Licht- und Sichtwegen sowie zum Merksatz 2. Die Begriffe Sicht- und Lichtwege sowie deren Konstruktion wurden vorab nochmals mit einem Tafelbild demonstriert.

9. Stunde: Das 1. Spiegelgesetz

Aus der 7. Stunde stand noch die Frage nach dem Ort der Spiegelbilder im Raum. Ihr sollte in dieser Unterrichtseinheit mit Schülerexperimenten nachgegangen werden. Die Schülerinnen und Schüler bearbeiteten in Zweier- oder Dreiergruppen das Aufgabenblatt B.5 (S. 203 im Anhang).

Für die Versuche wurden zwei Teelichter, eine Spiegelkachel mit Halter, Lineal und Bleistift benötigt. Im ersten Versuch sollte die Lage des Spiegels zwischen den vorgegebenen Orten der Kerze und deren Spiegelbild ermittelt werden. Das Ergebnis zeigte, dass der Spiegel senkrecht zur Verbindungslinie zwischen Kerze und Spiegelkerze stehen muss.

Im zweiten Versuch musste eine reale Kerze hinter dem Spiegel an den Ort gestellt werden, an dem das Spiegelbild der zweiten Kerze davor gesehen wurde. Das Einzeichnen der Schattengrenzen (siehe Übung zu Lichtwegen in der 8. Stunde, B.4 im Anhang) ermöglichte die genaue Bestimmung des Ortes.

Am Ende der Stunde wurden die Ergebnisse mit einem von den Schülerinnen und Schülern angefertigten Tafelbild verglichen und das 1. Spiegelgesetz formuliert:

Merksatz 3 (1. Spiegelgesetz): Gegenstand und Spiegelbild sind gleich weit von der Spiegelebene entfernt. Die Verbindungslinie zwischen Gegenstand und Spiegelbild steht senkrecht auf der Spiegelebene. (Mikelskis et al., 2006, S. 95)

10. Stunde: Die Spiegelwelt

Nachdem das Spiegelgesetz experimentell erarbeitet wurde, sollte die Spiegelwelt als eine optisch real wirkende Welt kennengelernt werden. In einem Demonstrationsexperiment war ein großer Spiegel (60 cm × 50 cm) senkrecht auf einem Tisch aufgebaut. Durch die genaue Positionierung in der Mitte des Tisches entstand der Eindruck, es handle sich um eine durchsichtige Scheibe, da die realen Kanten des Tisches mit den gespiegelten ineinander übergingen.



Abb. 8.2 Experimenteller Aufbau zur Spiegelwelt. Die Schatten, welche die Kerze an sich selbst, an der Spiegelkerze und umgekehrt verursacht, sind auf dem Foto nicht zu erkennen.

Hinter dem Spiegel wurde eine Kerze entzündet und der Raum ansonsten abgedunkelt. Anschließend wurde die Kerze langsam vor den Spiegel gestellt. Gleichzeitig wurden von den Schülerinnen und Schülern die Schattengrenzen an der Spiegelkante beobachtet, wie sie es bereits im Schülerexperiment getan haben. In dem Moment, in dem die Kerze die Spiegelebene durchbrach, schien sich die Situation der Schatten zu invertieren, was wiederum Erstaunen auslöste. Zusätzlich zur Kerze wurde eine schwarze Papprolle vor dem Spiegel gestellt (siehe Abb. 8.2)². Die Anordnung entsprach der Skizze auf Arbeitsblatt B.6 (S. 205 im Anhang), in dem die Schülerinnen und Schüler zunächst alle sichtbaren Schatten (insgesamt acht) einzeichnen sollten. Anschließend musste diesen Bereichen die zugehörige Lichtquelle und Schattengeber zugeordnet werden.

Bei der Bearbeitung dieses Arbeitsblattes war es erstmals nicht mehr möglich, sich physisch in die eingebundene Perspektive zu begeben. Dennoch konnten die Schülerinnen und Schüler sich gedanklich in die Position der Ameise versetzen und angeben, welcher Körper die Sicht zu welcher Kerze verdeckt.

Neben den Schatten, welche die Kerze an der Papprolle und an sich selbst verursachte, gab es auch solche, die von der Spiegelkerze an der Spiegelpapprolle und sich selbst herbeigeführt wurden. Außerdem schienen diese Schatten in den Spiegel hinein und aus dem Spiegel hinaus zu ragen (siehe Abb. 8.2 auf der rechten Seite). Eine weitere Art von Schatten waren jene, welche die Kerze an der Spiegelpapprolle und die Spiegelkerze an der Papprolle hervorbrachte. Letzterer ist in Abb. 8.2 deutlich als derjenige Schatten zu erkennen, der auf den Betrachter zeigt. Die Schattenbereiche konnten nun leicht mit denen aus dem Schülerexperiment (3. Stunde) in Verbindung gebracht werden. Es gab

²Die Anordnung des Versuchs geht auf Julius (1984) zurück.

Teilschatten, von denen nur eine der beiden Kerzen nicht sichtbar war und sogar einen Kernschatten ganz dicht am Spiegel (auf Abb. 8.2 rechts, wo sich der in die Spiegelwelt hineinragende Schatten mit dem aus ihr herausragenden Schatten kreuzt), von dem die Sicht zu beiden Kerzen von einer Papprolle verdeckt war.

Die Erkenntnis, die aus den Beobachtungen gewonnen wurde, lautete, dass die Spiegelbilder von der Kerze und der Papprolle nicht nur real *wirken*, weil sie den Anschein erwecken, als befänden sie sich tatsächlich hinter der Spiegelebene. Sie verursachen auch eine reale optische (*Aus-*)*Wirkung*. Die Spiegelbilder können Schatten in der Welt vor dem Spiegel hervorrufen, genauso wie von Objekten vor dem Spiegel Schatten an den Spiegelbildern verursachen können. In Bezug auf optische Effekte, hier am Beispiel der Schattenprojektion, ist die Welt hinter dem Spiegel (*Spiegelwelt*) real.

Mit dieser Erkenntnis konnten die Schülerinnen und Schüler nun endgültig das Doppelschattenexperiment verstehen. Als Konstruktionshilfe erhielten sie Arbeitsblatt B.7 (S. 206 im Anhang). Mithilfe des 1. Spiegelgesetzes konnten sie den Ort der Spiegel- leuchte und der Spiegelhand genau konstruieren. Außerdem waren sie nun fähig, alle Schatten einzuzeichnen, die von den beiden optisch realen Leuchten verursacht wurden. Die räumliche Lage von Leuchte, Schattengeber und Spiegel zueinander entsprach der Situation im Experiment zur Spiegelwelt. Der doppelte Schatten an der Zimmerdecke war damit nun kein Rätsel mehr.

11. Stunde: Das 2. Spiegelgesetz

Aufgrund der begrenzten Zeit war dies die letzte Stunde zum Thema *Spiegel* in dieser Klasse. In ihr sollte das Reflexionsgesetz am ebenen Spiegel (2. Spiegelgesetz) erarbeitet werden. In den anderen Versuchsklassen ergab sich noch die Möglichkeit einer anschließenden Unterrichtseinheit zum 3. Spiegelgesetz.

Nach einer Wiederholung der bisherigen Erkenntnisse zum 1. Spiegelgesetz und der Spiegelwelt fiel die Aufmerksamkeit wieder auf einen großen, vor der Klasse aufgebauten Spiegel (60 cm × 50 cm). Jeder konnte nun wenigstens ein Spiegelbild eines Klassenkameraden sehen. Beispielhaft wurden zwei Schülerinnen der Klasse (Anna und Sara)³ ausgewählt, die einander im Spiegel sehen konnten. Anna saß vorne links in der Klasse und Sara hinten rechts. Es wurde vom Lehrer die Frage formuliert, ob es nun eigentlich wichtig war, dass ein so großer Spiegel verwendet wurde, oder die beiden sich auch in einem kleineren Spiegel gesehen hätten. Dies wurde noch verschärft, indem beide einäugig in das offene Auge der Anderen schauen sollten. An dieser Stelle wurde vom Lehrer großer Wert darauf gelegt, dass hier von vier beteiligten Personen die Rede sein muss: Anna, Sara, Spiegel-Anna und Spiegel-Sara. Anna sah als Spiegel-Sara in das Auge und umgekehrt sah Sara Spiegel-Anna in das offene Auge (siehe Abb. 8.3).

Der Lehrer stellte sich hinter den Spiegel und ließ sich mit der Spitze eines Stifts von Anna an die Stelle des Spiegels bewegen, an der sie das Auge von Spiegel-Sara sah.

³Namen geändert

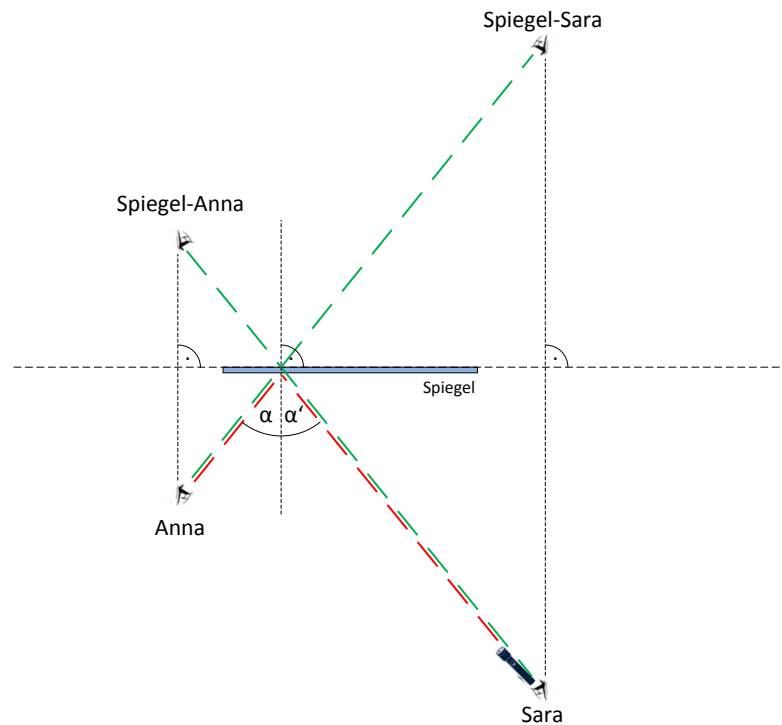


Abb. 8.3 Tafelbild zur Herleitung des 2. Spiegelgesetzes

Dort machte er eine Markierung. Schließlich sollte auch Sara den Lehrer mit seinem Stift zu der Stelle führen, an der sie Spiegel-Annas Auge sah. Dies war jedoch nicht notwendig, da sie anmerkte, dass es sich um die bereits markierte Stelle handelte. Die beiden Sichtwege kreuzten sich also genau auf der Spiegelebene, was in einem Tafelbild festgehalten wurde (Abb. 8.3, grün gestrichelt).

Nun sollte aber eine Aussage über die Lichtwege am Spiegel getroffen werden. Als eine Möglichkeit der Operationalisierung kannten die Schülerinnen und Schüler die Schattengrenzen, mit denen bisher Lichtwege sichtbar gemacht wurden. Sara erhielt eine Taschenlampe, mit der sie (Spiegel-)Anna blenden sollte. Ein dritter Schüler sollte mit einem Schattengeber den Weg von der Lampe bis zu Anna abgehen, sodass sie gerade nicht geblendet wurde. Anna half ihm dabei, den richtigen Weg zu finden. Dieser Grenzweg der Abschattung entsprach, wie sonst die Schattengrenzen, dem gesuchten Lichtweg (Abb. 8.3, rot gestrichelt). Nachdem das Tafelbild um die Lichtwege erweitert und die Beobachtung festgehalten wurde, dass der Lichtweg am Spiegel seine Richtung ändert, konnte über die Ähnlichkeit der Dreiecke in der Skizze das Reflexionsgesetz hergeleitet werden.

Merksatz 4 (2. Spiegelgesetz): Der Lichtweg ändert am Spiegel so seine Richtung, dass der Winkel zwischen einfallendem und reflektiertem Lichtbündel vom Lot auf die Spiegelebene halbiert wird: $\alpha = \alpha'$.

An dieser Stelle sei jedoch angemerkt, dass die Begriffe *Ähnlichkeit* und *Kongruenz* im Allgemeinen bei den Schülerinnen und Schülern der 7. Klassenstufe noch nicht bekannt sind. In diesem Fall konnte das Problem umgangen werden, da sie im Mathematikunterricht bereits Stufen-, Wechsel- und Scheitelwinkel kennengelernt hatten. So war die Herleitung von $\alpha = \alpha'$ über einen kleinen Umweg trotzdem möglich.

12. Stunde: Lernerfolgskontrolle

Als Abschluss der Themen *Licht und Schatten* und *Spiegelwelt* wurde eine schriftliche Lernerfolgskontrolle durchgeführt (siehe B.8 im Anhang). Sie beinhaltete Aufgaben zu Sicht- und Lichtwegen, zu Schattenbildern, zur Spiegelwelt und den Spiegelgesetzen.

13. Stunde: Experimente zur Hebung

Die letzten beiden Stunden, die für den Optikunterricht zur Verfügung standen, sollten der Einführung in das Phänomen der optischen Hebung dienen. Als Einstiegsexperiment wurde dafür der *geknickte Stab* demonstriert.

Ein rechteckiges, undurchsichtiges Kunststoffbecken ($60\text{ cm} \times 40\text{ cm} \times 32\text{ cm}$) wurde vor dem Unterricht randvoll mit Wasser gefüllt. In das Becken wurden zwei ca. 1,5 m lange und 2 cm breite Aluschienen gelegt. Eine der beiden Schienen wurde vorab genau am Auflagepunkt der Beckenkante so nach unten gebogen, dass beim schrägen Blick in das Wasser die Schiene gerade erscheint.

Nachdem die Schülerinnen und Schüler den Klassenraum betreten hatten, wurden sie gebeten, sich um das Becken herum zu stellen. Der Vergleich der beiden Schienen führte zu der Beobachtung, dass eine Schiene geknickt und eine gerade aussah. Schließlich wurden beide Schienen aus dem Wasser herausgeholt. Diejenige, die vorher gerade aussah, war nun geknickt und die Schiene, die geknickt aussah, war eigentlich gerade. Offensichtlich gab es beim Blick ins Wasser einen Konflikt zwischen der sichtbaren und der ertastbaren Erfahrung.

Mit der Motivation diesen Konflikt genauer zu untersuchen, wurden die Schülerinnen und Schüler in Experimentierstationen zur optischen Hebung geschickt. In kleinen Gruppen sollten sie nacheinander die Aufgabenblätter B.9 bearbeiten. In der Station *Harpunieren* reproduzierten sie die Beobachtung des Eingangsexperiments, dass die Gegenstände unter Wasser angehoben erscheinen. Hier hatten sie den Vorteil, dass sich jeder selbst davon überzeugen konnte, ob die Stange gerade war oder richtig auf den Stein gezielt wurde. In der Station *Senkrechte Hebung* wurde die Beobachtung schließlich quantifiziert. Die Wasserhöhe konnte von jedem selbst bestimmt werden, sodass am Ende in jeder Schülergruppe unterschiedliche Messwerte zur Verfügung standen.

14. Stunde: Die Brechung der Lichtwege

Am Anfang wurden die Ergebnisse der Schülerexperimente aus der vorherigen Stunde gesammelt und mit einem Tafelbild gesichert. Die Messwerte aus der letzten Station wurden für jede Gruppe notiert und die Quotienten a/b gemittelt. Folgende Beobachtungen konnten festgehalten werden:

- Der nach unten geknickte Stab sah beim Blick ins Wasser gerade aus.
- Der gerade Stab sah beim Blick ins Wasser nach oben geknickt aus.
- Das Zielfernrohr zielte auf den Stein, die Harpune traf aber darüber.
- Die Münze schien um $1/4$ der Wassertiefe angehoben.

Merksatz 5 (Optische Hebung): Beim Blick ins Wasser erscheinen uns die Dinge im Wasser zur Oberfläche hin lotrecht angehoben. Dies wird als *optische Hebung* bezeichnet. (Mikelskis et al., 2006, S. 102)

Das Vorgehen bis hierher war mit dem beim ebenen Spiegel vergleichbar. Zunächst wurde die Lage der gesehenen Objekte (Bilder) vermessen und anhand von Gesetzen festgelegt. Beim Spiegel war der Ort des Spiegelbildes noch unabhängig vom Betrachter und daher objektiv bestimmbar. Bei der Hebung traf dies zumindest beim senkrechten Blick auch zu, sodass eine quantitative Ortsangabe möglich war. Für den schrägen Blick war dies qualitativ in Relation zum realen Objekt möglich. Nun sollte wieder der Weg des Lichts untersucht werden, um zu prüfen, ob er, wie schon beim Spiegel, von den Sichtwegen abweicht. Um sich die Lichtwege am Übergang Luft-Wasser zu veranschaulichen, sollten wieder die Schattengrenzen als Werkzeug zur Hilfe genommen werden.

Ein mit Wasser gefülltes Aquarium wurde vor der Klasse gut sichtbar positioniert. Dabei musste beachtet werden, dass jeder auch frontal auf das Becken schaut und nicht durch die Seitenflächen des Aquariums. In das Becken wurde an der Rückseite eine große, matt weiße Kunststoffplatte schräg angestellt. Sie hatte etwa die Breite des Aquariums, war jedoch deutlich höher. Sie sollte als Projektionsfläche für die Schattengrenzen dienen. Die seitliche Oberkante des Beckens wurde mit undurchsichtigem Klebeband abgeklebt, was später als Schattengeber dienen sollte. Schließlich wurde eine helle Experimentierleuchte (Punktleuchte) angeschaltet, die seitlich neben der Kante mit dem Klebeband gestellt wurde. Ihr Schein streifte schräg nach unten die Kunststoffplatte.

Es waren gleichzeitig zwei Schattengrenzen sichtbar (Ober- und Unterkante des Klebebandes), die stellvertretend für zwei verschiedene Lichtwege standen. Beide schlossen mit der Wasseroberfläche unterschiedliche Winkel ein. An der Grenzfläche Luft-Wasser machten beiden Lichtwege einen Knick nach unten. Die Beobachtung wurde mit einem Tafelbild festgehalten und der letzte Merksatz der Unterrichtssequenz lautete:

Merksatz 6 (Brechung): Beim Eindringen des Lichts von Luft in Wasser macht der Weg des Lichtes an der Grenzfläche Luft-Wasser einen Knick hin zum Lot. Diese Richtungsänderung heißt *Brechung*. (Mikelskis et al., 2006, S. 104)

Schließlich musste noch die Verbindung zwischen den Phänomenen der Hebung und der Brechung der Lichtwege hergestellt werden. Dies konnte leider nicht zufriedenstellend beendet werden. Zunächst wurde die Umkehrbarkeit der Lichtwege thematisiert, die bisher noch nicht benötigt wurde. In einem zusätzlichen Experiment in der 1. Pilotstudie wurden dazu die Lichtquelle und der Schattengeber auch unter Wasser gestellt. Die Schattengrenzen inklusive der Abknickungen verliefen qualitativ gleich. Aber auch in diesem Fall musste für die weitere Argumentation der Sichtweg bei den Hebungsexperimenten mit der Schattengrenze über Wasser gleichgesetzt werden. Gleichzeitig war es nötig, plausibel zu machen, dass dieser Sichtweg gerade weiter verläuft. Um den Zusammenhang zwischen Brechung und Hebung zu verstehen, mussten also häufig die Ansichten der eingebundenen Perspektive mit denen der abgelösten in Einklang gebracht werden, was den Schülerinnen und Schülern sehr schwer viel. Aufgrund der Tatsache, dass keine weitere Physikstunde für die Intervention in der Hauptstudie zur Verfügung stand, wurde die phänomenbasierte Anfangsoptik mit diesem Stand beendet.

Im Vorfeld zur Hauptstudie fanden zwei auf das Projekt bezogene Pilotstudien statt, die im Folgenden vorgestellt werden sollen. Die in Abschnitt 2.2 erwähnte Vorstudie zu Identifizierung der Merkmale von PbPU fällt nicht darunter. Ihre Ergebnisse wurden bereits dort verarbeitet und Details zur Durchführung sind bei Westphal et al. (2011) dargestellt.

9.1 Erste Pilotierung

Die erste Pilotstudie hatte den Zweck, das Treatment und die Testinstrumente auf ihre Durchführbarkeit unter Realbedingungen zu prüfen. Außerdem sollten trotz der relativ kleinen Stichprobe die psychometrischen Eigenschaften der Tests abgeschätzt werden können.

Design und Stichprobe

Diese Erprobung war ähnlich wie das Design der Hauptstudie angelegt (Abb. 7.1). Im Gegensatz dazu wurde jedoch auf eine Kontrollgruppe und den Follow-Up-Test verzichtet. Die Stichprobe bestand aus $N = 32$ Schülerinnen und Schülern einer 7. Klasse an einem Berliner Gymnasium. Sie setzte sich aus 15 Jungen (47 %) und 17 Mädchen (53 %) zusammen.

Testinstrumente

Im Vor- und im Nachtest wurden die Assoziationen für das Image/Selbstkonzept, das Interesse und das Fachwissen erhoben. Für das Interesse und das Fachwissen kamen die

in Abschnitt 7.2 vorgestellten Fragebögen zum Einsatz. Die expliziten Assoziationen wurden mit einer Vorversion der semantischen Differentiale aus Abschnitt 7.2 gemessen.

Die Erhebung der impliziten Assoziationen war für die Hauptstudie ursprünglich mithilfe von ST-IATs geplant. Der Grund dafür war, dass in der Verwendung von nur einer Zielkategorie ein großer Vorteil gesehen wurde, weil durch die relative Messung bezüglich eines zweiten Faches eine unkontrollierte Störvariable existiert. Im Kontext von Vergleichs- und Veränderungsmessungen wäre es denkbar, dass sich während des Treatments in einer Gruppe die Assoziationen zu *Englisch* verändern, weil dort z.B. ein neues Thema begonnen wurde. Dies würde Einfluss auf die *D*-Werte haben. Daher wurden in dieser Pilotstudie die in Abschnitt 7.2 vorgestellten ST-IATs in den drei Dimensionen *Schwierigkeit*, *Geschlechtskonnotation* und *Selbstkonzept* erprobt.

Geräte und Prozedur

Vor dem Beginn der phänomenbasierten Unterrichtssequenz wurden die Teilnehmer an zwei Tagen getestet. Am ersten Testtag füllten alle zusammen in einer Unterrichtsstunde (45 min) den Fachwissensfragebogen im Klassenraum aus. Sie hatten dafür die kompletten 45 min Zeit, brauchten dafür jedoch durchschnittlich nur ca. 35 min. Am zweiten Testtag führte im Teilungsunterricht jeweils die Hälfte der Klasse die drei ST-IATs am Computer im Fachraum durch. Dies geschah an 15 tragbaren PCs (Lenovo N500, 2.00 GHz Intel Pentium Dual CPU) mit 39.1 cm Bildschirmdiagonale. Die Präsentation der Stimuli und die Aufzeichnung der Latenzzeiten geschah mit Inquisit 3.0.4.0 [Computer Software]. Der Text erschien weiß bzw. grün auf schwarzem Hintergrund. Die Zuordnung nach links geschah durch Druck der Taste *W*, die nach rechts durch Druck der Taste *P*. Der Ablauf der Tests entsprach der Darstellung in den Abschnitten 4.1.1, 4.2.1 und A.1.

Ebenfalls am PC wurden demografische Daten (Geschlecht und Computernutzung) sowie die expliziten Assoziationen erhoben. Die Dauer betrug dafür ca. 20 min. Im Anschluss daran wurden die Interessefragebögen ausgefüllt (ca. 10 min, Papier & Bleistift).

Die Intervention entsprach im Wesentlichen dem in Kapitel 8 vorgestellten Treatment. Die Klasse wurde in dieser Zeit durch den Autor unterrichtet. Bis auf wenige Ausnahmen war der eigentliche Physiklehrer der Klasse im Unterricht anwesend. Im Anschluss an die achtwöchige Intervention fanden die Nachtest analog den Vortests statt.

Ergebnisse und Diskussion

Als Ergebnis der ersten Pilotierung kann zunächst festgehalten werden, dass die Durchführung der Unterrichtssequenz zur phänomenbasierten Anfangsoptik ohne größere Probleme möglich war. Dies ist insofern nicht verwunderlich, als dass sie in dieser oder einer ähnlichen Form bereits von verschiedenen Lehrern erprobt wurde. Dennoch wurden die Feinplanung und der Unterricht selbst das erste Mal durch den Autor durchgeführt.

Die Erhebungen verliefen ebenfalls unproblematisch. Allerdings fiel auf, dass für die Durchführung der ST-IATs ein hohes Maß an Konzentration nötig ist. In einem Klassenverband von ca. 15 Schülerinnen und Schüler gibt es jedoch häufig Unruhe, auch aufgrund der ungewohnten Situation, sodass nicht bei jedem Teilnehmer der Test bis zum Ende mit Sorgfalt durchgeführt wurde. Bei den Papier & Bleistift-Fragebögen war dies kein Problem.

Der Interessefragebogen funktionierte zufriedenstellend bis gut. Die Skalenreliabilitäten liegen im Bereich $.72 \leq \alpha \leq .93$ und sind mit den Angaben im Skalenhandbuch vergleichbar (Frey et al., 2009). Zwischen Vor- und Nachtest gibt es keine signifikanten Unterschiede, was aber auf die kleine Stichprobengröße zurückzuführen ist. Dennoch sind auch die Effektstärken gering, bis verschwindend gering.

Der Fachwissenstest erwies sich ebenfalls als tauglich für die Hauptstudie. Ob die Schwierigkeit für die Zielpopulation angemessen war, ließ sich wegen der wenigen Teilnehmer, die zudem alle in einer Klasse waren, schwer einschätzen. Dafür ist die Varianz zu gering. Eine Rasch-Analyse wurde nicht vorgenommen. Trotzdem konnte ein hoch signifikanter Wissenszuwachs ermittelt werden, $t(28) = -4.37$, $p < .001$, $r = .82$. Das entspricht einem großen Effekt (siehe Tabelle C.1 im Anhang).

Die semantischen Differentiale erwiesen sich nur in der Dimension *Schwierigkeit* als problematisch, weil dort beide nicht miteinander korrelierten. Dafür war augenscheinlich eine ungünstige Formulierung verantwortlich, die für die Hauptstudie angepasst wurde (siehe Abschnitt 7.2).

Die mit dem ST-IAT gemessenen impliziten Assoziationen waren nicht zufriedenstellend. Die mittleren *D*-Werte waren alle (Vor- und Nachtest) verschwindend gering und zeigten somit nicht die erwarteten stereotyp-konformen Assoziationen auf. Da nur die Daten einer Klasse zur Verfügung standen, konnten wenig sinnvolle Gruppen für Vergleiche gebildet werden, um herauszufinden, ob der Test nicht sensibel genug ist oder ob im Mittel keine impliziten Assoziationen vorliegen. Zwischen den Geschlechtern gab es keine Unterschiede. Auch die Reliabilitäten waren nicht ausreichend, um weitere Aussagen über die Kennwerte zu treffen ($\alpha \leq .43$). Das führt zu dem Schluss, dass davon ausgegangen werden muss, dass mit den ST-IATs in der Zielpopulation keine zufriedenstellende Assoziationseffekte (*D*-Werte) gemessen werden. Schließlich musste für die Hauptstudie eine Alternative gesucht werden, worin die Motivation für die zweite Pilotstudie lag.

9.2 Zweite Pilotierung

Die zweite Pilotierung wurde im Rahmen einer Masterarbeit von Rehfeldt (2012) durchgeführt. Das Ziel bestand darin, drei verschiedene implizite Verfahren auf ihre Testgüte und vor allem auf die Stärke der Assoziationseffekte hin zu vergleichen. Der Test mit den besten Eigenschaften sollte schließlich in der Hauptstudie Anwendung finden.

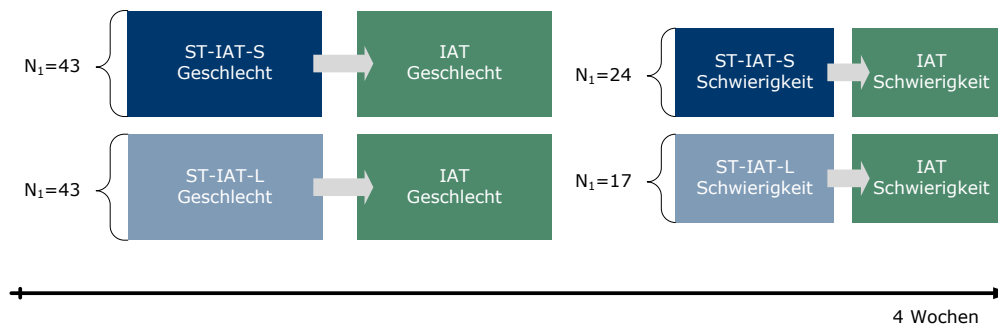


Abb. 9.1 Design der zweiten Pilotstudie (angelehnt an Rehfeldt, 2012, S. 38)

Design und Stichprobe

Die Studie war als Laboruntersuchung angelegt, in der jeder Teilnehmer zwei verschiedene Tests zur selben Dimension absolvierte (siehe Abb. 9.1). Auf die Dimension *Selbstkonzept* wurde verzichtet, weil dies erstens im Rahmen einer Masterarbeit zu aufwändig gewesen wäre und zweitens die untersuchten Hypothesen als von der IAT-Dimension unabhängig angenommen wurden.

Um zu prüfen, ob ein IAT höhere *D*-Werte als ein ST-IAT misst, absolvierte jeder Teilnehmer zunächst den ST-IAT und danach den IAT (Within-Subject). Da sich bei mehreren Tests hintereinander die Assoziationseffekte abschwächen, wurde der IAT immer als Zweites durchgeführt, da bei ihm höhere *D*-Werte angenommen wurden. Dies führt zu konservativen Entscheidungen.

Gleichzeitig sollte geprüft werden, ob sich die Reliabilität des ST-IATs durch mehr Trials verbessern lässt. Daher wurde in einem Between-Subject-Vergleich der in der ersten Pilotierung eingesetzte ST-IAT-S mit einem längeren Test verglichen (ST-IAT-L).

Die Stichprobe bestand aus insgesamt $N = 108$ Teilnehmern ($N_{\sigma} = 77$, $N_{\phi} = 31$) im Alter von 20 bis 71 Jahren, die mit überwiegendem Anteil Studierende der Physik waren. Damit ist sie nicht repräsentativ für die Absolutwerte der Assoziationsmaße und entstammt auch nicht aus der Zielpopulation der Hauptstudie. Da es sich jedoch um eine vergleichende Untersuchung handelt, in der nur die Tendenzen der Unterschiede für weitere Entscheidungen relevant sind, kann dieser Umstand entkräftet werden. Es gibt keine theoretischen Anhaltspunkte dafür, dass das Vorzeichen der Differenz zwischen den *D*-Werten von ST-IAT und IAT altersabhängig oder kontextspezifisch ist. Das Gleiche gilt für mögliche Unterschiede der Reliabilitäten von ST-IAT-L und ST-IAT-S.

Testinstrumente

IAT und ST-IAT-S entsprechen den Instrumenten, die in Abschnitt 7.2 vorgestellt wurden. Für den ST-IAT-L wurde die Anzahl der Trials im Testblock von 35 auf 65 Trials erhöht, die Stimuli blieben gleich. Dies entspricht noch immer einer geringeren Trialan-

zahl als von Karpinski und Steinman (2006) empfohlen wird. Jedoch ist sie fast doppelt so groß, wie die vom ST-IAT-S.

Neben den impliziten Assoziationen wurden demografische Daten (Geschlecht, Alter, Studiengang, Physik-Wahlverhalten in der Schule und Abiturdurchschnitt) und die expliziten Assoziationen mithilfe der entsprechenden semantischen Differentiale erhoben.

Geräte und Prozedur

Die Erhebungen fanden im Verlauf der vier Wochen unregelmäßig statt. Die Teilnehmer konnten an elf PCs gleichzeitig in einem Labor unter ruhigen Bedingungen die zwei impliziten Tests, die Befragung und die semantischen Differentiale hintereinander durchführen.

Wie schon in der ersten Pilotierung, wurden tragbare PCs (Lenovo N500, 2.00 GHz Intel Pentium Dual CPU) mit 39.1 cm Bildschirmdiagonale verwendet. Die Präsentation der Stimuli und die Aufzeichnung der Latenzzeiten erfolgte mit Inquisit 3.0.4.0 [Computer Software]. Der Text erschien weiß bzw. grün auf schwarzem Hintergrund. Die Zuordnung nach links geschah durch Druck der Taste *W*, die nach rechts durch Druck der Taste *P*. Der Ablauf der Tests entsprach der Darstellung in den Abschnitten 4.1.1, 4.2.1 und A.1.

Ergebnisse und Diskussion

Der für den Kontext der vorliegenden Arbeit entscheidende Schluss aus den Ergebnissen der zweiten Pilotstudie lautet, dass sich ein IAT besser für die Erhebung der impliziten Assoziationen zum Image eignet als ein ST-IAT.

In der Dimension *Geschlechtskonnotation* war der *D*-Wert beim IAT signifikant größer als beim ST-IAT-S, $t(42) = 2.94$, $p < .01$, $r = .41$ (mittlerer Effekt). Beim IAT wurden mittlere Assoziationen von *Physik* zu *männlich* gemessen ($M = 0.31$, $SD = 0.29$, $SE = 0.04$), während es beim ST-IAT-S nur geringe Assoziationen waren ($M = 0.15$, $SD = 0.30$, $SE = 0.05$). Zwischen dem IAT und dem ST-IAT-L ($M = 0.29$, $SD = 0.28$, $SE = 0.04$) gab es keine Unterschiede bezüglich des *D*-Wertes, $t(42) = -0.26$, $p = .79$. In der Dimension *Schwierigkeit* konnten keine signifikanten Unterschiede festgestellt werden.

Die Reliabilität konnte durch die Erhöhung der Trialanzahl nicht verbessert werden. Sie war beim ST-IAT-L ($.48 \leq \alpha \leq .74$) eher geringer als beim ST-IAT-S ($.74 \leq \alpha \leq .85$). Beim IAT hingegen war sie in allen Dimensionen zufriedenstellend und höher als bei den ST-IATs ($\alpha = .87$).

Die Überlegenheit des IATs in Bezug auf die Reliabilität entspricht den bisherigen Erfahrungen in anderen vergleichenden Studien und ist somit wie erwartet (siehe Abschnitt 4.2.2). Die Tatsache, dass durch eine Erhöhung der Trialanzahl die Reliabilität hier nicht größer wurde, unterstützt nicht die Aussagen von Karpinski und Steinman

(2006) und war somit nicht erwartungsgemäß. Eine Diskussion über mögliche Ursachen dafür gibt Rehfeldt (2012). Dies jedoch führt zu dem Schluss, dass der ST-IAT-L keine Alternative für die Hauptstudie ist, weil er zu keiner Verbesserung bezüglich der ersten Pilotstudie führen würde.

Eine Ursache für die niedrigeren *D*-Werte des ST-IATs gegenüber dem IAT könnte darin liegen, dass bei der Sortieraufgabe durch ein Ausschlussverfahren sortiert wird. Dafür müssten die Stimuli nur unter der Kategorie beurteilt werden, die allein steht. Fällt die Beurteilung positiv aus, wird die entsprechende Taste gedrückt, für alle anderen Stimuli, die jeweils andere Taste. Somit wäre es nicht mehr so sehr relevant, ob die beiden zusammenstehenden Kategorien kompatibel sind oder nicht, was den Assoziationseffekt abschwächen würde.

Da der IAT somit im Allgemeinen höhere Assoziationseffekte mit höherer Genauigkeit (Reliabilität) misst, als die beiden Alternativen, ist er das geeignete Messinstrument für die Hauptstudie.

Hauptstudie: Evaluation der phänomenbasierten Anfangsoptik

10.1 Beschreibung der Stichprobe

Die Hauptstudie fand an vier verschiedenen Berliner Gymnasien statt (Schule I-IV, Tabelle 10.1). Die Klassen der Versuchsgruppe wurden genauso wie die der Kontrollgruppe von vier verschiedenen Lehrern unterrichtet. Alle Lehrpersonen waren männlich. In der Versuchsgruppe waren dies drei Lehrer, die mit der phänomenbasierten Anfangsoptik sehr vertraut waren, weil sie aus dem nahen Umfeld der AG Didaktik der Physik an der Humboldt-Universität zu Berlin stammen, z. B. durch Studium, Mitarbeit im Uni-Lab Schülerlabor, Abordnung und/oder Mitarbeit in Projektgruppen (SINUS/PiKo). An ihrer Schule (I) wird sie bereits seit über zehn Jahren standardmäßig unterrichtet. Der vierte Lehrer der Versuchsgruppe war der Autor der vorliegenden Arbeit. Die vier Lehrer der Kontrollgruppe unterrichteten an drei verschiedenen Schulen. Sie kannten das untersuchte Optik-Curriculum nicht.

An Schule II waren zwei Klassen in der Versuchsgruppe (Lehrer NW) und zwei Klassen in der Kontrollgruppe (Lehrer RR und AN), sodass zumindest in diesen Klassen die Schule als Störvariable ausgeschlossen ist. Die Aufteilungen der Physikstunden unterschieden sich an den einzelnen Schulen, waren aber über beide Gruppen miteinander vergleichbar. An Schule I und IV wurde an zwei Stunden (45 min) pro Woche die ganze Klasse unterrichtet. An Schule II fand eine Stunde (45 min) pro Woche Teilungsunterricht statt. An Schule III wurden beide Stunden in der Woche nur in halben Klassen unterrichtet, was die geringe Anzahl N in Tabelle 10.1 erklärt. Im Durchschnitt hatte demnach die Kontrollgruppe mit 0.8 h mehr Teilungsunterricht pro Woche als die Versuchsgruppe (0.4 h). Dennoch sind bis auf jeweils eine Klasse pro Gruppe die anderen paarweise miteinander vergleichbar, sodass der Vorteil für die Kontrollgruppe gering ist.

Tabelle 10.1 Zusammensetzung der Stichprobe

Schule	Lehrer	Klasse	N	$N_{\text{♀}}$	$N_{\text{♂}}$	
I	AH	1	31	17	13	Versuchsgruppe
	PB	2	29	17	11	
	CG	3	30	15	12	
II	NW	4	32	18	14	
		7	31	21	10	
III	RR	5	32	17	15	Kontrollgruppe
	AN	6	33	22	11	
	GM	8	15	8	7	
		9	16	8	7	
IV	TK	10	30	17	13	
		11	30	15	15	
Summe VG	4	5	153	88	60	
Summe KG	4	6	156	87	68	

Anmerkung Die Anzahl an Teilnehmern pro Klasse N kann von der Summe der weiblichen und männlichen Teilnehmern abweichen, da von einigen das Geschlecht unbekannt ist .

Von den insgesamt $N = 309$ Teilnehmern sind 57 % weiblich ($N_{\text{♀}} = 175$) und 41 % männlich ($N_{\text{♂}} = 128$). Diese Verteilung ist in beiden Gruppen gleich, $\chi^2(1) = 0.34$, $p = .56$. Der optimale Umfang der Stichprobe wurde mit der Software G*Power 3.1.3 berechnet (Faul, Erdfelder, Lang & Buchner, 2007). Bei einem Signifikanzniveau von $\alpha = .05$ und einer Teststärke von $(1 - \beta) = 0.95$ können damit kleine Effekte aufgedeckt werden (Korrelationen, t -Tests, ANOVA-Interaktionen).

10.2 Datenerhebung

Die Erhebung der Daten und die Interventionen fanden je nach Schule zu unterschiedlichen Zeiten innerhalb von 10 Monaten statt. In jeder Klasse wurde unmittelbar vor Beginn des Optik-Moduls der Vortest an der jeweiligen Schule durchgeführt. Dafür wurden in der Regel zwei Testtage oder zwei Unterrichtsstunden am selben Tag benötigt. In einer Testeinheit wurde der Fachwissenstest (Papier & Bleistift, siehe Abschnitt 7.2) mit der gesamten Klasse durchgeführt. Die Bearbeitungszeit betrug ca. 35 min, konnte aber bei Bedarf die kompletten 45 min überdauern.

In der zweiten Testeinheit wurden an 15 tragbaren PCs¹ die drei IATs und die semantischen Differentiale durchgeführt. Die Reihenfolge der Dimensionen war für jeden Teilnehmer identisch: 1. *Schwierigkeit*, 2. *Selbstkonzept* und 3. *Geschlechtskonnotation*. Die Reihenfolge richtete sich nach der Relevanz im Sinne der Forschungsfrage. Aufgrund von Übungseffekten nehmen die Assoziationsmaße im Verlauf mehrerer hintereinander durchgeführter IATs ab (siehe Abschnitt 4.1.3). Da bei der Dimension *Geschlechtskonnotation* die geringsten Einflüsse von PbPU zu erwarten waren, wurde sie als Letztes getestet.

Die andere Hälfte der Gruppe bearbeitete gleichzeitig die Interessefragebögen (Papier & Bleistift). Nach ca. 20 min tauschten die Teilnehmer und führten den jeweils anderen Testteil der Einheit durch. Die Reihenfolge der Testeinheiten richtete sich nach schulorganisatorischen Gegebenheiten und konnte demnach variieren. Die Erhebungen der Nachtests fanden analog zu den Vortests statt. In der Regel waren die zeitlichen und räumlichen Bedingungen für jede Klasse im Vor- und Nachtest identisch.

Nachdem die Daten des Vor- und Nachtests ausgewertet waren, zeichnete sich bereits ab, dass bei den meisten abhängigen Variablen zwischen Versuchs- und Kontrollgruppe keine von der Testzeit abhängigen Unterschiede existieren. In Bezug auf die Forschungsfrage wäre es demnach nicht sinnvoll gewesen, den geplanten Follow-Up-Test trotzdem durchzuführen, da dort potenziell auftretende Unterschiede nicht mehr mit dem Treatment in Verbindung gebracht werden könnten. Auch aus Gründen der Testethik wäre es nicht vertretbar gewesen, die Versuchsteilnehmer (Schülerinnen und Schüler, Lehrer sowie die Schulen) ein weiteres Mal zu belästigen. Daher wurden nur Vor- und Nachtest erhoben und ausgewertet.

10.3 Subjektive Eindrücke von den Erhebungen

Im Folgenden werden einige der subjektiven Eindrücke zusammengefasst, die bei den Erhebungen gewonnen werden konnten. Der Autor der vorliegenden Arbeit war als Versuchsleiter bei jeder Testeinheit anwesend.

Die äußeren Bedingungen für die Durchführung der Tests waren an den verschiedenen Schulen sehr unterschiedlich. Sie hatten jedoch m. E. großen Einfluss auf die Ruhe und Konzentration bei der Bearbeitung der IATs. Aufgrund der Anzahl an zur Verfügung stehenden tragbaren PCs einschließlich der Softwarelizenzen (Inquisit 3.0.4.0 [Computer Software]) musste in der Regel die Klasse aufgeteilt werden. In den Fällen, wo mit der jeweiligen Hälfte der Raum nicht gewechselt werden konnte, führte dies zu starker Ablenkung durch die Teilnehmer, die bereits mit der Bearbeitung ihres Fragebogens fertig waren und auf einen Wechsel warteten. Ebenfalls nachteilig waren die zum Teil beengten Platzverhältnisse, die bei einem Wechsel zweier Teilnehmer zu weiteren Störungen führten. Hinzu kommt, dass die Probanden mitunter sehr dicht nebeneinander

¹Lenovo N500, 2.00 GHz Intel Pentium Dual CPU mit 39.1 cm Bildschirmdiagonale.

sitzen mussten und somit durch den Bildschirm des Nachbarn ebenfalls abgelenkt waren.

Als wichtiger Faktor kann außerdem die Anwesenheit des Klassenlehrers genannt werden, die nicht immer möglich war. Besonders problematisch fiel dies bei den Erhebungen des Vortests in den Klassen 1, 2 und 3 sowie beim Nachtest in Klasse 4 auf. Die große Unruhe und geringe Motivation war bereits bei der Durchführung daran zu erkennen, dass die IATs zum Teil nicht sorgfältig beendet wurden. Ein vorzeitiger Abbruch der Tests war nicht möglich. Er konnte nur beschleunigt werden, indem die Tasten *W* und *P* schnell nacheinander gedrückt wurden.

Auffällig war, dass bei der Bearbeitung der schriftlichen Tests und Fragebögen kaum Probleme der Konzentration und Sorgfalt auftraten. Insgesamt erweckt es also den Eindruck, dass die Bedingungen in einer Feldstudie für die Durchführung von IATs nicht optimal sind, weil sie sehr stark von der Aufmerksamkeit und Motivation der Probanden abhängen. Die Tatsache, dass nicht für jeden Teilnehmer ein PC zur Verfügung stand, schien die geschilderte Problematik zu verstärken.

10.4 Datenaufbereitung, Testgüte und Skalierung

Implizite Assoziationen

Die Aufbereitung der Daten der drei IATs erfolgte nach dem Algorithmus zur Berechnung der *D*-Werte (siehe Abschnitt 4.1.2 und Tabelle 4.1). Er spezifiziert auch die Löschung von Trials und von Fällen. Trials mit einer Latenzzeit von mehr als 10 s wurden nicht berücksichtigt. Dort kann davon ausgegangen werden, dass der Proband abgelenkt war. Der Rest des Tests kann jedoch, davon unabhängig, weiter ausgewertet werden.

Für jeden Test wurden die Fälle gelöscht, die in mehr als 10 % der Trials Latenzzeiten aufwiesen, die kleiner als 300 ms waren. Diese Probanden haben einen zu großen Anteil des Tests nicht gewissenhaft durchgeführt, sondern die Tasten *W* und *P* schnell abwechselnd gedrückt, um ein vorzeitiges Ende des Tests zu erreichen. Die Löschung wurde testweise vorgenommen, da die Teilnehmer mitunter nur eine Dimension „abgebrochen“ haben. Somit konnte in den anderen Dimensionen die Fallzahl höher gehalten werden. Im Vortest wurden auf diese Weise in der ersten Dimension *Schwierigkeit* 3 % (VG 4 %, KG 1 %), in der zweiten Dimension *Selbstkonzept* 8 % (VG 9 %, KG 6 %) und in der dritten Dimension (*Geschlechtskonnotation*) 9 % (VG 12 %, KG 7 %) gelöscht. Im Nachtest waren es 10 % (VG 8 %, KG 14 %), 18 % (VG 15 %, KG 20 %) und 23 % (VG 24 %, KG 22 %; gleiche Reihenfolge). Hier ist bereits zu erkennen, dass die Abbrecherquote mit zunehmender Anzahl an absolvierten IATs zunimmt und die Testmotivation im Nachtest geringer war, als im Vortest.

Die teilweise hohen Quoten an aussortierten Fällen, führen aufgrund der Messwiederholung zu geringerer Teststärke bzw. Sensitivität, da für jeden Teilnehmer das Vor- und Nachtestergebnis benötigt wird. Bei ungünstigen Konstellationen kann der Prozent-

satz an nicht berücksichtigten Fällen noch größer werden, wenn die aussortierten Fälle des Vortests nicht mit denen des Nachtests übereinstimmen.

Die restliche Prozedur entspricht der Darstellung in Abschnitt 4.1.2 (siehe auch Anhang A.1). Die Ergebnisse der Datenaufbereitung der IATs sind für jeden Teilnehmer insgesamt sechs *D*-Werte, aus drei Dimensionen je einer im Vor- und im Nachtest. Ein positiver *D*-Wert entspricht einer stereotyp-kompatiblen Assoziation von *Physik* mit *schwierig*, *Junge* oder *Andere*.

Für die Schätzung der Reliabilität der IATs wurde für jeden Stimulus die Differenz der Latenzzeit zwischen kompatiblen und inkompatiblen Block gebildet. Damit durch die hohe Trialanzahl die interne Konsistenz nicht künstlich zu groß werden lässt, wurden über fünf Differenzen Mittelwerte gebildet. Für jeden Teilnehmer ergab das 12 mittlere Latenzzeitunterschiede pro IAT. Aus ihnen wurde die Reliabilität aus der internen Konsistenz (Cronbachs α) abgeschätzt. Die Werte liegen im Bereich $.72 \leq \alpha \leq .93$ und sind damit für implizite Verfahren mehr als zufriedenstellend (siehe Tabelle C.2 im Anhang).

Explizite Assoziationen

Die zwei jeweils zu einer Dimension gehörigen semantischen Differentiale korrelieren alle höchst signifikant mit großen Effekten: Schwierigkeit $\rho = .60$, $p < .001$; Geschlechtskonnotation $\rho = .76$, $p < .001$; Selbstkonzept $\rho = .86$, $p < .001$.² Somit ist eine Mittelwertbildung zwischen beiden Items zulässig. Dafür wurden die Ergebnisse der negativ gepolten Differentiale transformiert, sodass die Werte zwischen 5 und 10 einer stereotyp-kompatiblen Assoziation von *Physik* mit *schwierig*, *Junge* oder *Andere* bedeutet ($5 \hat{=}$ neutral). Aus der Mittelwertbildung resultieren für jeden Teilnehmer insgesamt sechs explizite Assoziationsmaße, in drei Dimensionen je einer im Vor- und im Nachtest.

Interesse

Die Reliabilitäten für die Interessesskalen sind bis auf die Subskala *wau*³ ($\alpha_{\text{Nachtest}} = .60$) mit Cronbachs $\alpha > .70$ zufriedenstellend (siehe Tabelle C.3). Da bei Persönlichkeitstests die Werte für die Reliabilität oft geringer ausfallen, als z. b. in Leistungstests oder als die in der psychodiagnostischen Literatur oft geforderte untere Grenze von $\alpha \geq .80$, werden hier Werte mit $\alpha \geq .70$ als zufriedenstellend eingeschätzt (vgl. Rost, 2005). Hinzu kommt, dass die Instrumente für Gruppenvergleiche eingesetzt werden und somit die jeweiligen Gruppenmittelwerte weitaus zuverlässiger sind, als die Einzelwerte jeder Person.

²Da die semantischen Differentiale streng genommen ordinalskaliert und die Werte zusätzlich nicht normalverteilt sind, wurde das nichtparametrische Korrelationsmaß Spearmans ρ verwendet.

³für die Zuordnung der Kürzel zur vollständigen Skalenbezeichnung und Beispielimens siehe Tabelle 7.1, S. 104

Die statistischen Kennwerte für das Interesse an Physikunterricht und seinen Inhalten ergeben sich aus den einzelnen Skalenmittelwerten für jeden Teilnehmer. Eine hohe Zahl steht für eine positive Ausprägung in der jeweiligen Skala.

Fachwissen

Die Daten des dichotom codierten Fachwissenstests wurden mithilfe einer Raschanalyse für die weiteren statistischen Auswertungen aufbereitet. Dafür wurde Winsteps 1.0.0 [Computer Software] verwendet. Die Itemschwierigkeiten, die von dem ESNaS-Projekt bekannt waren, konnten leider nicht ohne Weiteres übernommen werden, da die Zielpopulation der Normierung dort eine andere war (Ende Sek.I, Walpuski et al., 2010). Daher wurden die Schwierigkeiten im Vortest mit berechnet. Anhand dieser Parameter war es außerdem möglich, die Testgüte abzuschätzen.

Als Gütekriterien für den Test werden im Rahmen der probabilistischen Testtheorie die Item-Fit-Statistiken herangezogen. Sie geben an, inwiefern die Items den Annahmen des Modells entsprechen. Als akzeptabel für die Modellpassung werden hier Werte für die mittlere quadratische Abweichung mit $\text{Outfit-MNSQ} < 1.3$ und dem dazugehörigen Signifikanzniveau mit $-2 < t < 2$ ($p > .045$) angenommen (Bond & Fox, 2007). Sind für alle Items des Tests diese Kriterien erfüllt, gilt er als rasch-homogen. Das bedeutet insbesondere, dass die interne Konsistenz der Items⁴ und die spezifische Objektivität nachgewiesen ist. Die Ergebnisse hängen also nicht von der gewählten Stichprobe an Items oder Personen ab, was wiederum bedeutet, dass die Rasch-Homogenität auch für sämtliche Teilstichproben gilt (Kauertz, 2008).

Von den elf Items des Fachwissenstests erfüllt eines die Anforderungen nicht (siehe Tabelle C.4 im Anhang, S. 212). Für die weiteren Analysen und Schätzungen der Personenparameter wurde dieses Item nicht berücksichtigt. Eine zusätzliche Überprüfung der Testgüte stellt die Split-Half-Analyse dar. Bei ihr werden die Itemschwierigkeiten anhand zwei zufälliger Teilstichproben bestimmt und gegeneinander aufgetragen. Wenn die oben erwähnte Unabhängigkeit zutrifft, müssten die Punkte der Schwierigkeiten bei der Teilmengen auf einer Geraden liegen. In Abb. C.1 (Anhang) ist dieser lineare Zusammenhang für den Fachwissenstest gezeigt.

Anhand dieser beiden Kriterien (Fit-Statistiken und Split-Half-Analyse) konnten die verbliebenen zehn Items als rasch-homogen identifiziert werden und so mithilfe der Rasch-Analyse die Personenfähigkeiten geschätzt werden. Für den Nachtest wurden die Itemschwierigkeiten des Vortests als feste Parameter gesetzt.

Die Reliabilität des Tests ist nicht zufriedenstellend. Die den klassischen Maßen äquivalente Personen-Reliabilität liegt zwischen .41 (Real) und .48 (Modell). Das klassisch berechnete Maß liegt bei $KR_8 = .51$.⁵ Die bei der Rasch-Analyse geschätzte Perso-

⁴inklusive der Annahme, dass der Beantwortung der Items ein gemeinsames latentes Konstrukt zugrunde liegt

⁵Kuder-Richardson-Formel 8 (siehe Formel (C.1) im Anhang). KR_8 berücksichtigt im Gegensatz zu KR_{20} auch unterschiedliche Trennschärfen der Items (Lienert & Raatz, 1998).

nenreliabilität hängt laut Linacre (2014) stark von der Stichprobenvarianz, der Länge des Tests, der Anzahl an Kategorien pro Item und von der Passung zwischen Itemschwierigkeiten und Personenfähigkeiten ab. Mit der geringen Reliabilität ist auch eine ungenaue Schätzung der Personenparameter verbunden.

Ein Vorteil des Rasch-Modells liegt darin, dass Personenfähigkeit und Itemschwierigkeit auf derselben Skala abgebildet werden können. In Abb. C.2 werden die Verteilungen beider Parameter des Vortests miteinander verglichen. Je mehr sie sich überdecken, desto besser ist die Schwierigkeit des Tests auf die Fähigkeiten der Stichprobe angepasst. An der (geringen) Verschiebung der Personenparameter-Verteilung nach rechts erkennt man, dass der Test etwas zu leicht für die Teilnehmer war.

Die durch die Rasch-Analyse ermittelten Personenfähigkeiten stellen für die weiteren Auswerteverfahren die statistischen Kennwerte des Fachwissens dar.

Im folgenden Kapitel werden die Ergebnisse der Hauptstudie vorgestellt. Wie in Abschnitt 7.1 begründet wurde, werden für die Beantwortung der Forschungsfragen zweifaktorielle Varianzanalysen mit Messwiederholung auf einem Faktor durchgeführt. In den Fällen, bei denen die Vorannahmen des Verfahrens nicht verletzt sind, werden zusätzlich Kovarianzanalysen der Nachteststatistiken durchgeführt, mit den Vortestergebnissen als Kovariate. Zunächst wird in Abschnitt 11.1 die Überprüfung der Vorannahmen dargelegt. Folgend werden, untergliedert nach Testinstrumenten in Abschnitt 11.2 bis 11.5, die Ergebnisse der ANOVAs und ANCOVAs präsentiert, die der Hypothesenüberprüfung dienen. Daran schließt sich Abschnitt 11.6 an, mit den Ergebnissen von weiterführenden (explorativen) Untersuchungen.

11.1 Überprüfung der Vorannahmen

Da es sich bei der Varianzanalyse um ein parametrisches Verfahren handelt, gelten die gleichen Voraussetzungen wie für die meisten Methoden, die auf der Normalverteilung basieren. Für alle varianzanalytischen Vergleiche werden folgende Vorannahmen getroffen:

1. Unabhängigkeit der Zwischensubjektfaktoren
2. Intervallskaliertheit der Daten
3. Normalverteilung des Merkmals (in jeder Gruppe)
4. Varianzhomogenität der Zwischensubjektfaktoren

Bei einer Varianzanalyse mit Messwiederholung ist die Unabhängigkeit der Faktoren zwangsläufig nicht gegeben. An ihre Stelle tritt die

5. Sphärizität (Varianzhomogenität der Differenzen zwischen zwei messwiederholten Faktorstufen)

Sie ist vergleichbar mit der Varianzhomogenität bei den Zwischensubjektfaktoren und ein weniger strenger Spezialfall einer CS-Struktur der Kovarianzmatrix (Eid et al., 2011). Die Sphärizität ist bei nur zwei Messzeitpunkten (2 Stufen) immer gegeben (Field, 2009). Für die Kovarianzanalysen gibt es zwei zusätzliche Vorannahmen:

6. Unabhängigkeit der Kovariate vom Zwischensubjektfaktor
7. Homogenität der Regressionsanstiege

Die erste und grundlegendste Bedingung, die Unabhängigkeit der Daten zwischen den Teilnehmern muss im Rahmen dieser Studie ungeprüft vorausgesetzt werden. Sie wird insofern abgeschwächt, als dass im Rahmen des hier durchgeführten Mittelwertvergleichs zwischen zwei Gruppen, diese mit hoher Wahrscheinlichkeit nicht voneinander abhängig sind. Lediglich in den Klassen von Schule II waren sich Schülerinnen und Schüler beider Gruppen überhaupt bekannt. Der Rest ging auf Schulen in verschiedenen Stadtteilen. Eine Möglichkeit, die Abhängigkeit der Messwerte an einer Schule zu beachten, wäre die in Abschnitt 7.1 angesprochene Mehrebenenanalyse.

Die Intervallskaliertheit ist für die meisten Daten uneingeschränkt gegeben. Die *D*-Werte sind von Natur aus normierte Zeitunterschiede und damit auch metrisch. Das Gleiche trifft auf die durch die Rasch-Analyse gewonnenen Personenparameter zu. Die Interessesskalen und die semantischen Differentiale sind streng genommen nur ordinalskaliert. Bei den Differentialen kann dies vernachlässigt werden, da durch den Schieberregler eine quasi-stufenlose Positionierung möglich war. Ebenso ist es bei Likert-Skalen (Interesse) üblich, für jeden Teilnehmer Mittelwerte über die Skala zu bilden, die dann wiederum ein metrisches Maß sind.

Normalverteilung

Bis auf einen Fall (Vortest, Geschlecht, KG) sind alle *D*-Werte normalverteilt (Tabelle C.5 im Anhang). Die Mittelwerte der semantischen Differentiale und der Interessesskalen sowie die Personenfähigkeiten der Fachwissenstests sind es jedoch im Allgemeinen laut Kolmogorov-Smirnov-Test und Shapiro-Wilk-Test nicht (siehe den Tabellen C.6, C.7 und C.9 im Anhang)¹. Diese Abweichung kommt bei größeren Stichproben sehr

¹Die Tests überprüfen die Nullhypothese, dass die erhobene Verteilung nicht von der Normalverteilung abweicht. Bei signifikanten Testergebnissen (fett hervorgehoben) muss man also davon ausgehen, dass der Unterschied zwischen Normalverteilung und erhobener Verteilung nicht durch Zufall gefunden wurde.

häufig vor, da durch sie auch kleinste Unterschiede von der Normalverteilung signifikant werden. Gleichzeitig gilt jedoch, dass die Varianzanalyse bei großen Stichproben ($N > 30$) relativ robust gegenüber der Verletzung der Normalverteilungsannahme ist, vor allem, wenn die Stichproben in etwa gleich groß sind (Eid et al., 2011; Glass, Peckham & Sanders, 1972). Damit ist eine Durchführung der parametrischen Verfahren dennoch zulässig. Robuste bzw. nichtparametrische Verfahren, welche äquivalent zur zweifaktoriellen Varianzanalyse mit Messwiederholung auf einem Faktor und zur Kovarianzanalyse testen, existieren nicht (Field, 2009).

Varianzhomogenität

Die Überprüfung der Varianzhomogenität erfolgte durch den Levene-Test. Sie ist in allen Dimensionen bei den D -Werten und den Mittelwerten der semantischen Differentiale sowie bei den Personenparametern des Fachwissenstests im Vor- und Nachtest gegeben (siehe Tabellen C.10, C.11 und C.12 im Anhang).

Bei der Auswertung der Mittelwerte der Interessesskalen ist das Ergebnis des Tests in zwei Fällen signifikant (siehe Tabelle C.13 im Anhang) und somit dort die Annahme der Varianzhomogenität nicht zulässig. Allerdings trifft auf den Levene-Test das Gleiche wie auf die Normalverteilungstests oben zu. Bei großen Stichproben kann das Ergebnis signifikant werden, auch wenn die Abweichungen der Varianzen nur sehr gering sind. Im vorliegenden Fall muss jedoch auch nach der Methode des kritischen Varianzverhältnisses (Hartleys F_{Max}) von Varianzheterogenität ausgegangen werden.

In solchen Fällen kann der F -Test der ANOVAs jedoch abgeschätzt werden. Wenn die Varianz in einer kleineren Stichprobe *größer* ist als in einer größeren Stichprobe ($\hat{\sigma}_1^2 < \hat{\sigma}_2^2$ bei $n_1 > n_2$), dann entscheidet der F -Test zu liberal und die Wahrscheinlichkeit eines α -Fehlers erhöht sich. Wenn die Varianz in einer kleineren Stichprobe jedoch *kleiner* ist als in der größeren Stichprobe ($\hat{\sigma}_1^2 > \hat{\sigma}_2^2$ bei $n_1 > n_2$), dann entscheidet der F -Test zu konservativ. Die Wahrscheinlichkeit des α -Fehlers ist in diesen Fällen geringer als das nominelle α -Niveau (Eid et al., 2011; Glass et al., 1972).

Bei den beiden hier vorliegenden Mittelwerten der Interessesskalen ist die Varianz in der Kontrollgruppe geringer als in der Versuchsgruppe ($\hat{\sigma}_{KG}^2 < \hat{\sigma}_{VG}^2$), während das Verhältnis bei der Stichprobengröße umgekehrt ist ($N_{KG} > N_{VG}$). Somit neigen die folgenden varianzanalytischen Verfahren in diesen beiden Fällen zu liberalen Entscheidungen.

Unabhängigkeit der Kovariate

Für die Durchführung der Kovarianzanalysen dürfen sich die zu vergleichenden Gruppen (Zwischensubjektfaktor) keine gemeinsame Varianz der Kovariate teilen. Obwohl dies oft nicht beachtet wird, kann die ANCOVA keine Unterschiede bzgl. der Kovariaten „kontrollieren“ oder „ausbalancieren“ (Field, 2009; Miller & Chapman, 2001). Bei

den angestrebten ANCOVAs sollen die Nachtestergebnisse von Versuchs- und Kontrollgruppe miteinander verglichen und die Vortestergebnisse als Kovariate berücksichtigt werden. Es ist also für jede Variable zu überprüfen, ob sich die beiden Gruppen im Vortest unterscheiden.

Der Vergleich wurde mithilfe von T -Tests durchgeführt (siehe Tabelle C.14 im Anhang). Die Vorannahme ist in sieben Fällen verletzt: für die Mittelwerte der semantischen Differentiale in der Dimension *Schwierigkeit* ($t(261) = -3.540$, $p < .001$, $r = .21$) und *Selbstkonzept* ($t(261) = -2.998$, $p = .003$, $r = .18$), für die Mittelwerte der Interessesskalen *wku* ($t(258) = 2.763$, $p = .006$, $r = .17$), *wsewkuwau* ($t(263) = 2.317$, $p = .021$, $r = .14$), *mo3* ($t(231.7) = 2.349$, $p = .020$, $r = .15$) und *mo123* ($t(261) = 2.449$, $p = .015$, $r = .15$) sowie für die Personenparameter des Fachwissenstests ($t(261) = -2.154$, $p = .032$, $r = .13$).

Homogenität der Regressionsanstiege

Die Vorannahme über die Homogenität der Regressionsanstiege bedeutet anschaulich, dass im Zwischensubjektfaktor, also in beiden Gruppen, der Zusammenhang zwischen abhängiger Variable (Nachtest) und Kovariate (Vortest) gleich stark sein soll. Zur Überprüfung wurden für die Variablen, bei denen die Unabhängigkeit der Kovariate (siehe oben) gegeben war, ein allgemeines lineares Modell (GLM III, univariat) gerechnet und die Interaktion zwischen Kovariate und Gruppe ausgewertet (siehe Tabelle C.15 im Anhang). Ist der F -Test für diese Interaktion signifikant, sind die Regressionsanstiege in beiden Gruppen nicht gleich. Dies trifft nur auf den Mittelwert der Skala *mo1* zu, $F(1, 236) = 0.789$, $p = .45$, sodass auch diese Variable von den ANCOVAs ausgeschlossen werden muss.

11.2 Implizite Assoziationen

Die deskriptiven Statistiken der D -Werte im Vor- und Nachtest sind in Tabelle C.16 (Anhang) zusammengefasst. Die mittleren D -Werte unterscheiden sich zwischen beiden Gruppen in allen Dimensionen im Vortest nicht (siehe Tabelle C.14 im Anhang). In den Dimensionen *Geschlechtskonnotation* (VG $M = 0.09$, $SE = 0.04$; KG $M = 0.12$, $SE = 0.03$) und *Selbstkonzept* (VG $M = 0.04$, $SE = 0.03$; KG $M = 0.01$, $SE = 0.02$) liegen sie unterhalb einer schwachen Assoziation ($|D| > 0.15$, siehe Tabelle C.1 im Anhang). Im Vortest wird also in beiden Gruppen *Physik* weder mit *Junge/Mädchen* noch mit *Selbst/Andere* assoziiert. In der Dimension *Schwierigkeit* liegt in beiden Gruppen eine schwache Assoziation mit *schwierig* vor (VG $M = 0.18$, $SE = 0.03$; KG $M = 0.18$, $SE = 0.02$).

Die Nachtestergebnisse zeigen ein ähnliches Bild. Die Dimensionen *Geschlechtskonnotation* (VG $M = 0.12$, $SE = 0.03$; KG $M = 0.15$, $SE = 0.03$) und *Selbstkonzept* (VG $M = 0.06$, $SE = 0.03$; KG $M = 0.05$, $SE = 0.03$) weisen verschwindend gerin-

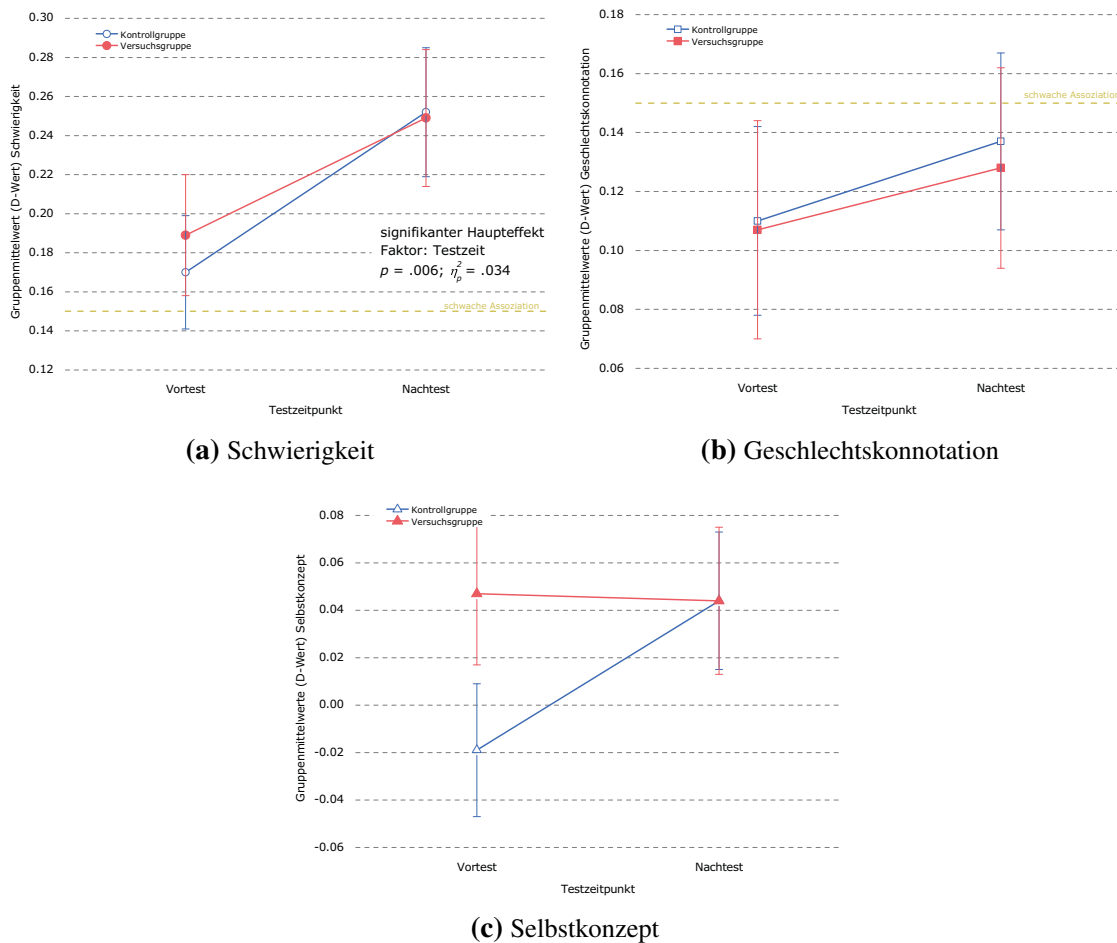


Abb. 11.1 Interaktionsdiagramme der drei IAT-Dimensionen *Schwierigkeit*, *Geschlechtskonnotation* und *Selbstkonzept*. Die Mittelwerte weichen geringfügig von den deskriptiven Statistiken ab, weil hier nur die Teilnehmer berücksichtigt sind, für die Vor- und Nachtestergebnisse vorliegen. Die Fehlerbalken stellen den Standardfehler *SE* dar.

ge Assoziationen auf, während bei *Schwierigkeit* in beiden Gruppen die *D*-Werte größer werden, jedoch im Bereich einer schwachen Assoziation bleiben (VG $M = 0.26$, $SE = 0.03$; KG $M = .026$, $SE = 0.03$).

Die zweifaktoriellen ANOVAs mit Messwiederholung ergeben nur für *Schwierigkeit* einen signifikanten Haupteffekt des Faktors *Testzeit* mit geringer Stärke, $F(1, 217) = 7.70$, $p = .006$, $\eta_p^2 = .03$ (siehe C.21). Die impliziten Assoziationen von *Physik* mit *schwierig* nehmen im Mittel über beide Gruppen hinweg zwischen den Testzeitpunkten zu (siehe Abb.11.1a). Ein signifikanter Haupteffekt des Faktors Gruppe liegt nicht vor. In den Dimensionen *Geschlechtskonnotation* und *Selbstkonzept* treten keine signifikanten Haupteffekte auf (Abb. 11.1b und 11.1c).

Die Interaktion zwischen dem Testzeitpunkt (Vor-/Nachtest) und der Gruppe (VG/KG) ist bei keiner Dimension signifikant (siehe Tabelle C.21 im Anhang). Die Effekte sind

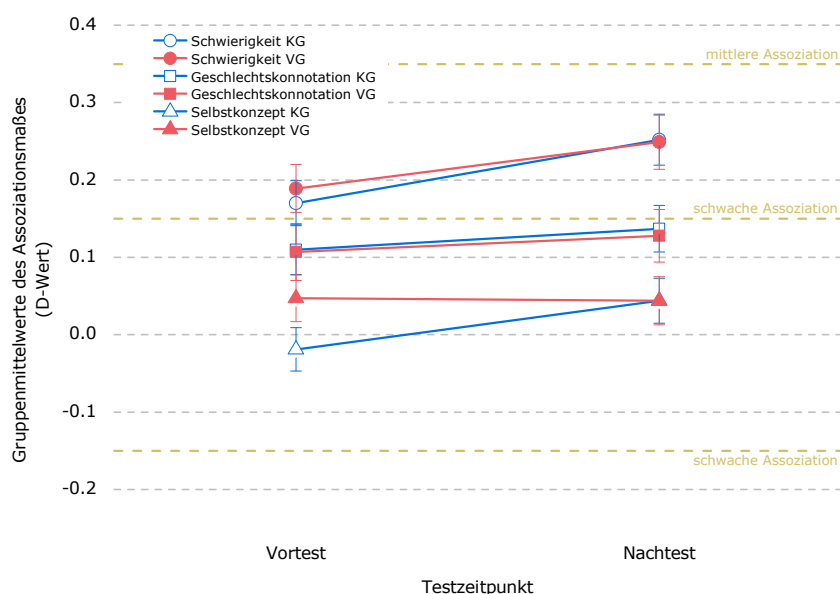


Abb. 11.2 Interaktionsdiagramme aller IAT-Dimensionen. Ein positiver D-Wert steht für eine Assoziation von *Physik* mit *schwierig*, *Junge* und *Andere*. Die Fehlerbalken stellen den Standardfehler *SE* dar.

verschwindend gering, $0,000 \leq \eta_p^2 \leq 0,007$. Betrachtet man die Trends qualitativ, so entwickeln sie sich alle bis auf eine Ausnahme zwischen den Testzeitpunkten hin zu Assoziationen von *Physik* mit *Junge* und *Andere* (siehe Abb. 11.2). Beim *Selbstkonzept* ist der Unterschied zwischen Vor- und Nachtest in der Kontrollgruppe marginal signifikant, $t(105) = -1,72$, $p = .09$, $r = .17$. Hier gibt es also einen kleinen Effekt bei der Verschiebung der Assoziation von *Ich* zu *Andere*, während die Assoziation in der Versuchsgruppe konstant bleibt (siehe Abb. 11.1c).

Die Ergebnisse der ANCOVAs stützen die der ANOVAs (siehe Tabelle C.26 im Anhang). In keiner der drei Dimensionen gibt es einen signifikanten Effekt auf dem Faktor *Gruppe*: Schwierigkeit $F(1,215) = 0.17$, $p = .68$, $\eta_p^2 = .001$; Geschlechtskonnotation $F(1,180) = 0.03$, $p = .87$, $\eta_p^2 = .000$; Selbstkonzept $F(1,193) = 0.08$, $p = .77$, $\eta_p^2 = .000$.

Die Post-Hoc-Teststärke für die Interaktionen der zweifaktoriellen ANOVAs für empirisch relevante, mittlere Effekte liegt bei $(1 - \beta) \geq .99$. Die Sensitivität bei einer geforderten Teststärke von $(1 - \beta) = .95$ war hoch genug, um mittlere Effekte aufzudecken. Die beobachtete Teststärke für die (beobachteten) Effektstärken liegt im Bereich $.05 < (1 - \beta) < .23$ (siehe Tabelle C.29 im Anhang).

Zusammenfassend kann festgehalten werden, dass die eingesetzten IATs nur schwache bis verschwindend geringe Assoziationen zu *Physik* messen. Zwischen Versuchs- und Kontrollgruppe gibt es keine Unterschiede, auch nicht in Abhängigkeit vom Test-

zeitpunkt. Die schwache Assoziation von *Physik* mit *schwierig* hat sich im Verlauf der Hauptstudie in beiden Gruppen gleichermaßen verstärkt.

11.3 Explizite Assoziationen

Die deskriptiven Statistiken der semantischen Differentiale sind in Tabelle C.17 (Anhang) zusammengefasst. In der Dimension *Geschlechtskonnotation* (VG $M = 6.16$, $SE = 0.16$; KG $M = 6.44$, $SE = 0.14$) unterscheiden sich beide Gruppen im Vortest nicht. *Physik* wird leicht mit *Junge* assoziiert. In den Dimensionen *Schwierigkeit* (VG $M = 4.49$, $SE = 0.18$; KG $M = 5.36$, $SE = 0.17$) und *Selbstkonzept* (VG $M = 4.20$, $SE = 0.23$; KG $M = 5.19$, $SE = 0.23$) gibt es schwache, aber signifikante Unterschiede zwischen beiden Gruppen im Vortest (siehe Abschnitt 11.1 und Tabelle C.14 im Anhang), jedoch sind die Assoziationen in beiden Dimensionen nahezu neutral.

Im Nachtest ist die Situation identisch. Bei *Schwierigkeit* (VG $M = 4.68$, $SE = 0.18$; KG $M = 5.28$, $SE = 0.19$) und *Selbstkonzept* (VG $M = 4.83$, $SE = 0.21$; KG $M = 5.53$, $SE = 0.21$) gibt es Gruppenunterschiede, jedoch im Bereich neutraler Assoziationen. Die schwache Assoziation zu *Junge* (VG $M = 6.28$, $SE = 0.17$; KG $M = 6.60$, $SE = 0.15$) bleibt ohne Unterschiede zwischen den Gruppen bestehen.

Die rein deskriptiven Aussagen werden durch die zweifaktoriellen ANOVAs mit Messwiederholung bestätigt (siehe Tabelle C.22 im Anhang und Abb. 11.3). Die Gruppenunterschiede bei *Schwierigkeit* und *Selbstkonzept* zeigen sich durch signifikante schwache Haupteffekte bei diesem Faktor, $F(1, 238) = 11.77$, $p = .001$, $\eta_p^2 = .05$ und $F(1, 238) = 8.28$, $p = .004$, $\eta_p^2 = .03$. In der Dimension *Selbstkonzept* gibt es außerdem einen signifikanten mittleren Haupteffekt bei dem Faktor *Testzeit*, $F(1, 238) = 15.25$, $p < .001$, $\eta_p^2 = .06$. Über beide Gruppen hinweg verändert sich die physikbezogene Assoziation in Richtung *Andere*, was eine Verschlechterung des physikbezogenen Selbstkonzepts bedeutet.

Signifikante Interaktionen zwischen beiden Faktoren gibt es nicht. Die Effekte sind verschwindend gering ($\eta_p^2 < .01$). Dieser Befund wird auch durch das Ergebnis der ANCOVA in der Dimension *Geschlechtskonnotation* gestützt.² Die expliziten gedanklichen Verknüpfungen zu *Physik* und deren Veränderung zwischen beiden Testzeitpunkten sind also unabhängig von der Zugehörigkeit zur Versuchs- bzw. Kontrollgruppe.

Die impliziten und expliziten Assoziationen zum Image von Physik (Schwierigkeit und Geschlechtskonnotation) korrelieren weder im Vor- noch im Nachtest miteinander. Die Effekte sind empirisch unbedeutend und hier auch nicht signifikant ($|r| < .10$). Im Unterschied dazu sind beim Selbstkonzept schwache Korrelationen vorhanden (Vortest $r = .17$, $p = .006$; Nachtest $r = .17$, $p = .009$).

Die Unterschiede zwischen den impliziten und expliziten Verfahren zeigen sich auch beim direkten Vergleich der Polaritätsprofile (Abb. 11.4). Im rechten Profil sind die

²Für die anderen Dimensionen sind die Vorannahmen der ANCOVA verletzt (Abschnitt 11.1).

11 Ergebnisse

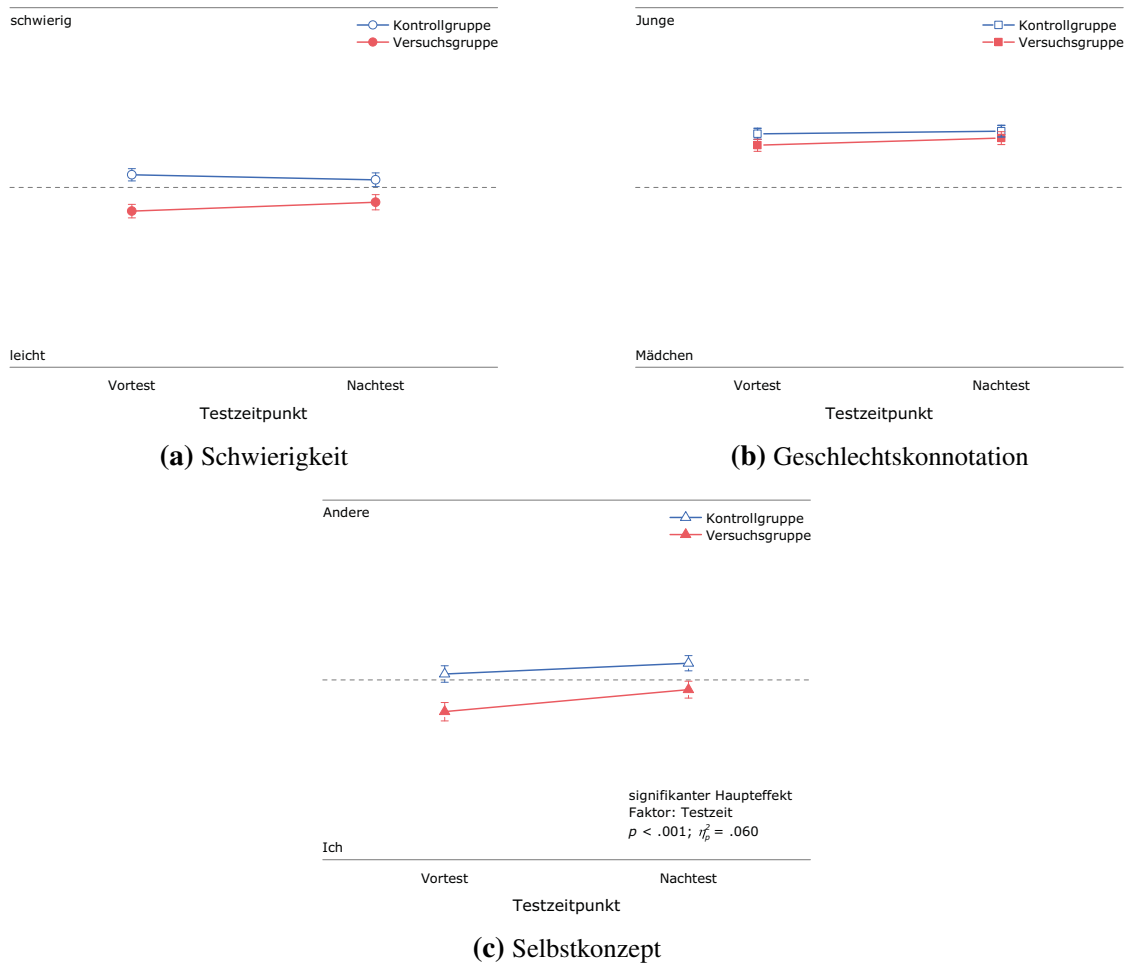


Abb. 11.3 Interaktionsdiagramme der expliziten Assoziationen in den Dimensionen *Schwierigkeit*, *Geschlechtskonnotation* und *Selbstkonzept*. Die Mittelwerte weichen geringfügig von den deskriptiven Statistiken ab, weil hier nur die Teilnehmer berücksichtigt sind, für die Vor- und Nachtestergebnisse vorliegen. Die Fehlerbalken stellen den Standardfehler *SE* dar.

Mittelwerte der semantischen Differentiale so dargestellt, dass deren Ort den Positionen des Schiebereglers entspricht. Zum Vergleich wurden die Mittelwerte der *D*-Werte so skaliert, dass eine starke implizite Assoziation einer maximalen Positionierung des Schiebereglers in den Differentialen entspricht. Es zeigt sich zunächst bei beiden Instrumenten, dass die Verschiebungen des Profils vom Vor- zum Nachtest in beiden Gruppen weitestgehend parallel verlaufen. Lediglich beim Selbstkonzept im IAT zeigt sich diesbezüglich ein qualitativer Unterschied zwischen den Gruppen (Abb. 11.4, links unten). Die Ausprägung der Assoziation unterscheidet sich bei beiden Instrumenten vor allem in der Dimension *Schwierigkeit*. Während der IAT schwache Assoziationen misst, sind die Selbstauskünfte der Schülerinnen und Schüler neutral. Der Unterschied zwischen den impliziten und expliziten Assoziationen, vor allem bei der Schwierigkeit und dem

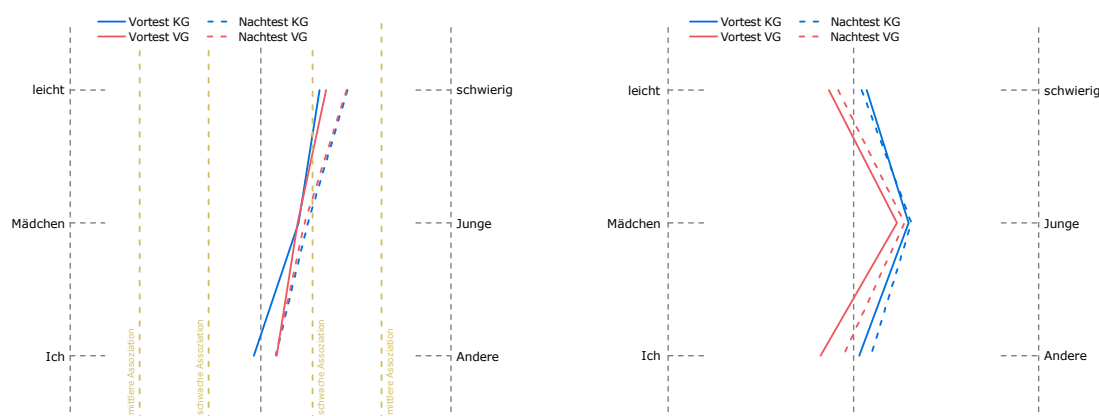


Abb. 11.4 Vergleich der Polaritätsprofile der impliziten (links) und der expliziten Assoziationen (rechts). Beide Diagramme wurden für den Vergleich so skaliert, dass eine starke Assoziation im IAT der maximal möglichen Angabe im semantischen Differential entspricht.

Selbstkonzept, ist augenscheinlich für die Versuchsgruppe größer als für die Kontrollgruppe.

Die Post-Hoc-Teststärke für die Interaktionen der zweifaktoriellen ANOVAs für empirisch relevante, mittlere Effekte liegt bei $(1 - \beta) \geq .99$. Die Sensitivität bei einer geforderten Teststärke von $(1 - \beta) = .95$ war hoch genug, um schwache Effekte aufzudecken. Die beobachtete Teststärke für die (beobachteten) Effektstärken liegt im Bereich $.09 < (1 - \beta) < .33$ (siehe Tabelle C.29 im Anhang).

Die semantischen Differentiale messen wie die IATs nur schwache bis verschwindend geringe Assoziationen. Sie zeigen keine Unterschiede zwischen Versuchs- und Kontrollgruppe, die vom Testzeitpunkt abhängig sind. Eine Veränderung der (expliziten) Assoziationen, unabhängig von der Gruppe, ist nur beim Selbstkonzept beobachtbar. Dies hat sich zwischen Vor- und Nachtest verschlechtert. Zwischen IATs und semantischen Differentialen gibt es bei den Image-Dimensionen keinen korrelativen Zusammenhang. Bei der Dimension *Schwierigkeit* sind auch die Ausprägungen der Assoziationen unterschiedlich.

11.4 Interesse

Die deskriptiven Statistiken aller Interessesskalen im Vor- und Nachtest sind in den Tabellen C.19 und C.20 (Anhang) zusammengefasst. Die Statistiken der (Ko-)Varianzanalysen sind in den Tabellen C.23, C.24 und C.28 (Anhang) aufgeführt.

Bei der Skala *Unterstützende Lehr-/Lernbedingungen* (wahrgenommene soziale Eingebundenheit, Kompetenzunterstützung und Autonomieunterstützung, siehe Abb. 11.5) unterscheiden sich im Vortest die beiden Gruppen signifikant, $t(263) = 2.317$, $p = .021$, $r = .14$. Die Ausprägung der wahrgenommenen Unterstützung ist in der Versuchsgruppe

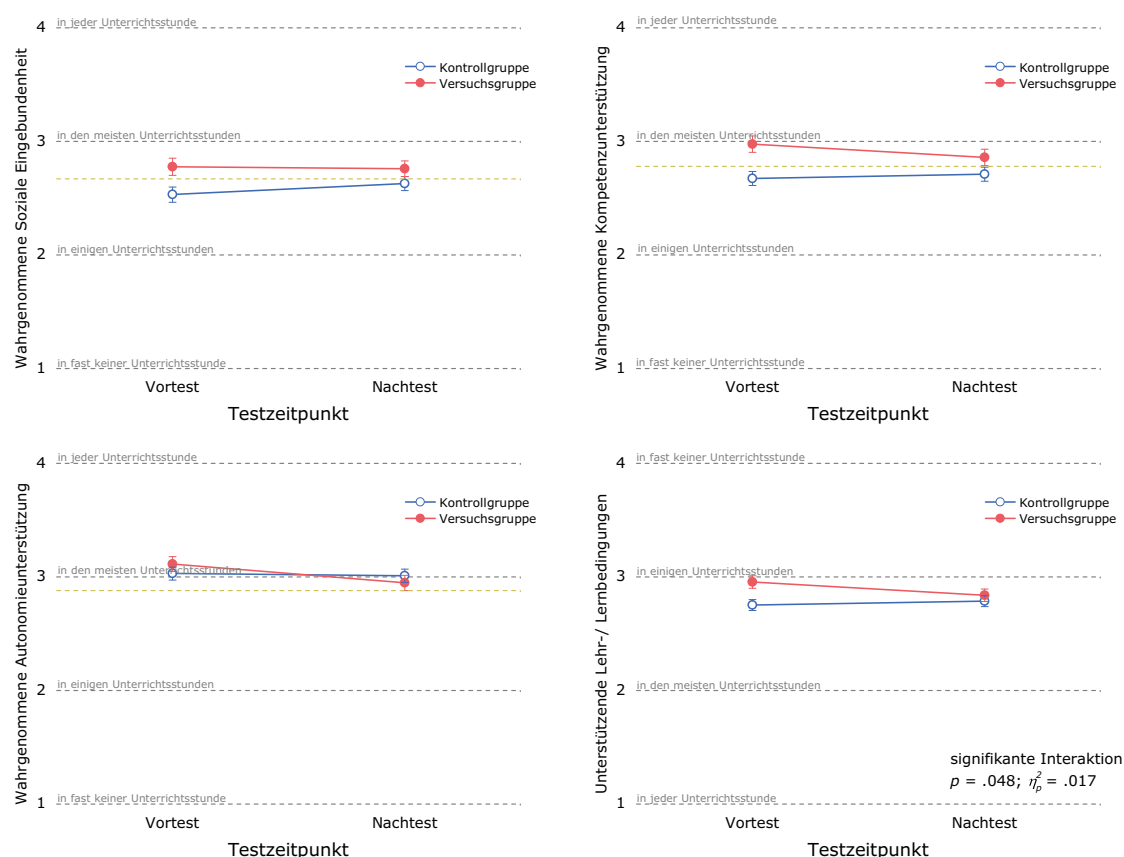


Abb. 11.5 Interaktionsdiagramme der Skala *Unterstützende Lehr-/Lernbedingungen* (u. r.) mit den drei Subskalen *Wahrgenommene Soziale Eingebundenheit* (o. l.), *Wahrgenommene Kompetenzunterstützung* (o. r.) und *Wahrgenommene Autonomieunterstützung* (u. l.). Die Fehlerbalken stellen den Standardfehler *SE* dar. Die gelbe gestrichelte Linie markiert den Skalenmittelwert von PISA 2006 (Frey et al., 2009).

($M = 2.91$, $SE = 0.06$) zunächst höher als in der Kontrollgruppe ($M = 2.74$, $SE = 0.05$). Dieser Unterschied ist auf die Subskala *wahrgenommene Kompetenzunterstützung* zurückzuführen (siehe Abschnitt 11.1 und Tabelle C.14 im Anhang). Bei der ANOVA weist der Faktor *Gruppe* einen schwachen, signifikanten Haupteffekt auf, $F(1, 233) = 4.05$, $p = .045$, $\eta_p^2 = .02$, d. h. unabhängig vom Testzeitpunkt unterscheiden sich Versuchs- und Kontrollgruppe. Außerdem existiert eine schwache Interaktion beider Faktoren *Testzeit* \times *Gruppe*, $F(1, 233) = 3.97$, $p = .048$, $\eta_p^2 = .02$. Die wahrgenommene Unterstützung der motivational günstigen Bedingungen sinkt also über die Zeit in der Versuchsgruppe (Nachtest $M = 2.84$, $SE = 0.05$) relativ zur Kontrollgruppe (Nachtest $M = 2.77$, $SE = 0.05$). Eine Untersuchung auf Subskalenebene bringt keine signifikanten Effekte hervor, die differenziertere Aussagen zulassen würden. Letzteres wird auch durch die nicht signifikanten ANCOVAs der Skalen *wse* und *wau* gestützt (siehe Tabelle C.28 im Anhang).

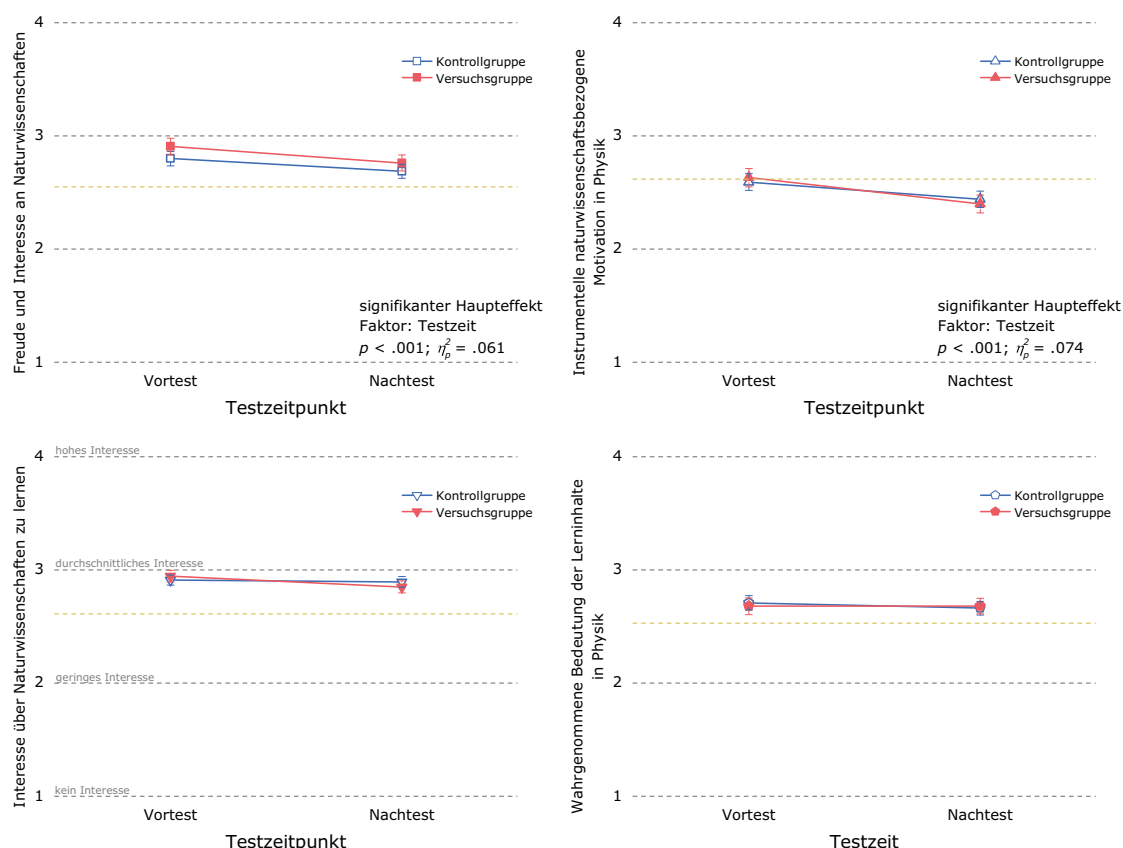


Abb. 11.6 Interaktionsdiagramme der Skalen *Freude und Interesse an Naturwissenschaften* (o.l.), *Instrumentelle naturwissenschaftsbezogene Motivation in Physik* (o.r.), *Interesse über Naturwissenschaften zu lernen* (u.l.) und *Wahrgenommene Bedeutung der Lerninhalte in Physik* (u.r.). Die Fehlerbalken stellen den Standardfehler *SE* dar. Die gelbe gestrichelte Linie markiert den Skalenmittelwert von PISA 2006 (Frey et al., 2009).

Bei der Skala *Freude und das Interesse an Naturwissenschaften* (Abb. 11.6 o.l.) treten weder im Vortest (VG $M = 2.87$, $SE = 0.7$; KG $M = 2.80$, $SE = 0.06$) noch im Nachtest (VG $M = 2.77$, $SE = 0.06$; KG $M = 2.68$, $SE = 0.06$) Gruppenunterschiede auf. Die zweifaktorielle ANOVA zeigt einen mittleren Haupteffekt (*Testzeit*), $F(1, 231) = 15.04$, $p < .001$, $\eta_p^2 = .06$. In beiden Gruppen sinkt die Freude und das Interesse also gleichermaßen. Die Werte bleiben aber noch oberhalb des Skalenmittelwerts von PISA 2006 (Frey et al., 2009).

Die instrumentelle physikbezogene Motivation (*INSTSCIE*, Abb. 11.6 o.r.) ist ebenfalls im Vortest in beiden Gruppen gleich hoch, (VG $M = 2.65$, $SE = 0.08$; KG $M = 2.56$, $SE = 0.06$), und sinkt zwischen den beiden Testzeitpunkten gleichermaßen stark (VG $M = 2.50$, $SE = 0.07$; KG $M = 2.43$, $SE = 0.07$). Dies drückt sich auch durch einen mittleren Haupteffekt (*Testzeit*) aus, $F(1, 215) = 17.20$, $p < .001$, $\eta_p^2 = .07$. Obwohl hier keine Varianzhomogenität angenommen werden darf und der *F*-Test zu liberal tes-

11 Ergebnisse

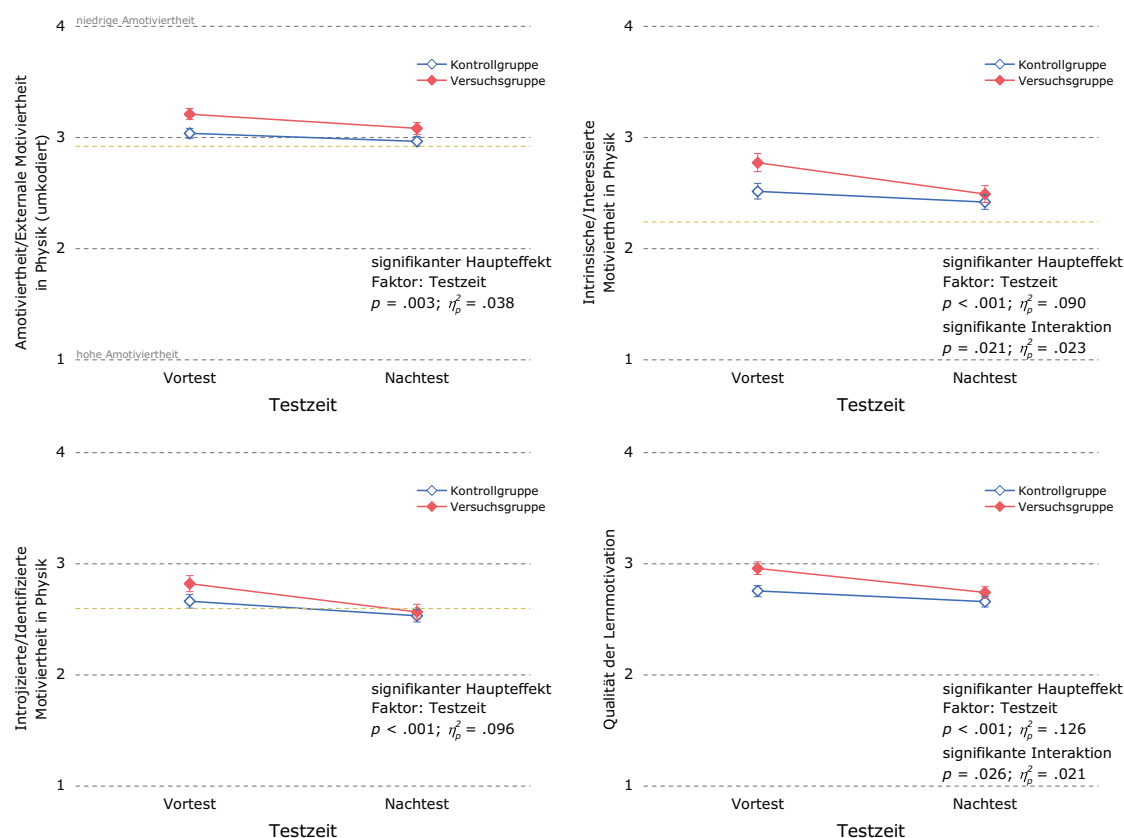


Abb. 11.7 Interaktionsdiagramme der Skala *Qualität der Lernmotivation* (u. r.) mit den drei Subskalen *Amotiviertheit/Externale Motiviertheit in Physik* (o. l.), *Intrinsische/Interessierte Motiviertheit in Physik* (o. r.) und *Introjizierte/Identifizierte Motiviertheit in Physik* (u. l.). Die Fehlerbalken stellen den Standardfehler *SE* dar. Die gelbe gestrichelte Linie markiert den Skalenmittelwert von PISA 2006 (Frey et al., 2009).

tet (siehe Kapitel 11.1), kann aufgrund der sehr hohen Signifikanz davon ausgegangen werden, dass der aufgetretene Effekt systematisch und nicht zufällig ist.

Das *Interesse über Naturwissenschaften zu lernen* (INTSCIE, VG $M = 2.95$, $SE = 0.05$; KG $M = 2.92$, $SE = 0.04$, Abb. 11.6 u. l.) und die *wahrgenommene Bedeutung der Lerninhalte in Physik* (meco, VG $M = 2.69$, $SE = 0.08$; KG $M = 2.72$, $SE = 0.06$, Abb. 11.6 u. r.) ist in beiden Gruppen gleich stark ausgeprägt und verändert sich im Laufe der Zeit nicht. Die ANOVAs und ANCOVAs zeigen keine signifikanten Haupt- und Interaktionseffekte. Die Ergebnisse der ANCOVAs der vier vorab betrachteten Skalen sind ebenfalls nicht signifikant und führen somit zu den gleichen Schlussfolgerungen.

Die *Qualität der Lernmotivation* (Abb. 11.7 u. r.) ist im Vortest in der Versuchsgruppe signifikant höher als in der Kontrollgruppe (VG $M = 2.94$, $SE = 0.05$; KG $M = 2.77$, $SE = 0.05$; $t(261) = 2.449$, $p = .015$, $r = .15$) und sinkt in beiden Gruppen in der Zeit bis zum Nachtest (VG $M = 2.77$, $SE = 0.04$; KG $M = 2.65$, $SE = 0.05$). Dies zeigt sich auch durch einen mittleren bis starken Haupteffekt der ANOVA, $F(1, 232) = 33.44$,

$p < .001$, $\eta_p^2 = .13$. Außerdem sinkt sie in der Versuchsgruppe stärker als in der Kontrollgruppe, $F(1, 232) = 5.03$, $p = .026$, $\eta_p^2 = .02$ (schwacher Effekt).

Bei einer differenzierten Betrachtung zeigen die Varianzanalysen, dass der Haupteffekt des Faktors *Testzeit* in allen Subskalen auftritt (siehe Abb. 11.7). Die signifikante Interaktion kann jedoch auf die Skala *Intrinsische/Interessierte Motiviertheit* („Im Unterricht in Physik bin ich mit Freude dabei“) zurückgeführt werden. Die intrinsische Motiviertheit sinkt dementsprechend in der Versuchsgruppe stärker als in der Kontrollgruppe, $F(1, 230) = 5.37$, $p = .021$, $\eta_p^2 = .02$. Bei dieser Skala ist die Vorannahme der Varianzhomogenität verletzt und aufgrund des Verhältnisses der Stichprobengrößen zu dem der Varianzen ist der F -Test zu liberal (siehe 11.1). Dennoch kann aufgrund der relativ hohen Signifikanz und des qualitativen Verlaufs der Linien im Interaktionsdiagramm (Abb. 11.7 o. r.) von einem nicht nur zufälligen Effekt ausgegangen werden.

Die Post-Hoc-Teststärke für die Interaktionen der zweifaktoriellen ANOVAs für empirisch relevante, mittlere Effekte bezüglich der Interessesskalen liegt bei $(1 - \beta) \geq .99$. Die Sensitivität bei einer geforderten Teststärke von $(1 - \beta) = .95$ war hoch genug, um mittlere Effekte aufzudecken. Die beobachtete Teststärke für die (beobachteten) Effektstärken liegt im Bereich $.08 < (1 - \beta) < .64$ (siehe Tabelle C.29 im Anhang).

Die Skalen zum Interesse an Physikunterricht und seinen Inhalten zeigen ein einheitliches Bild. Im Allgemeinen ist die absolute Ausprägung relativ hoch und oft höher als der Mittelwert von PISA 2006 (Frey et al., 2009). Die Tendenz ist allerdings in allen Skalen negativ. In den Fällen, in denen sich die Entwicklung in Versuchs- und Kontrollgruppe unterscheidet, nimmt das Interesse in der Versuchsgruppe stets stärker ab.

11.5 Fachwissen

Die deskriptiven Statistiken des Fachwissenstests sind in Tabelle C.18 (Anhang) zusammengefasst. Die Personenparameter des Vortests (VG $M = 0.62$, $SE = 0.11$; KG $M = 0.97$, $SE = 0.11$) sind bereits im Vortest im positiven Bereich, was darauf hindeutet, dass der Test für die Stichprobe etwas zu leicht war. Der Mittelwert der Itemschwierigkeiten ist auf 0 normiert. Die Versuchs- und Kontrollgruppe unterscheiden sich im Vortest signifikant, jedoch nur im geringen Ausmaß (siehe Abschnitt 11.1). Der Unterschied bleibt im Nachtest bestehen (VG $M = 1.16$, $SE = 0.12$; KG $M = 1.58$, $SE = 0.11$).

Die zweifaktorielle ANOVA mit Messwiederholung zeigt zwei signifikante Haupteffekte auf den Faktoren *Testzeit*, $F(1, 259) = 46.09$, $p < .001$, $\eta_p^2 = .15$ (starker Effekt) und *Gruppe*, $F(1, 259) = 7.62$, $p = .006$, $\eta_p^2 = .03$ (schwacher Effekt). Über beide Gruppen hinweg besteht also ein großer Zuwachs an Fachwissen und über beide Testzeitpunkte gemittelt, unterscheiden sich die beiden Gruppen im geringen Maße. Die Kontrollgruppe hat im Mittel ein höheres Fachwissen als die Versuchsgruppe. Eine signifikante Interaktion gibt es nicht, d. h. es können keine von der Testzeit abhängigen Gruppenunterschiede festgestellt werden (siehe Abb. 11.8).

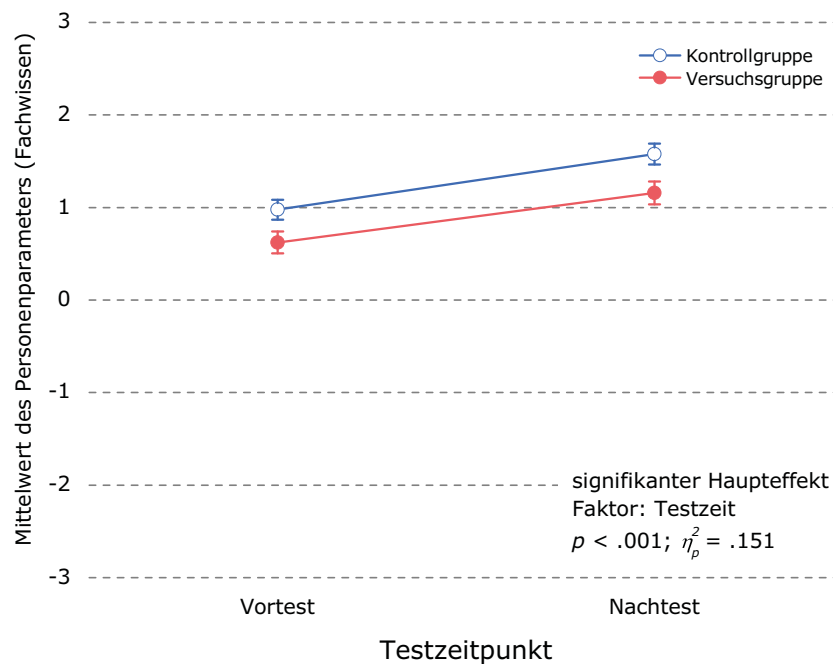


Abb. 11.8 Interaktionsdiagramme der Personenparameter nach der Raschanalyse des Fachwissenstests. Die Fehlerbalken stellen den Standardfehler *SE* dar.

Die Post-Hoc-Teststärke für die Interaktionen der zweifaktoriellen ANOVAs für empirisch relevante, mittlere Effekte liegt für den Fachwissenstest bei $(1 - \beta) \geq .99$. Die Sensitivität bei einer geforderten Teststärke von $(1 - \beta) = .95$ war hoch genug, um mittlere Effekte aufzudecken. Die beobachtete Teststärke für die (beobachtete) Effektstärke liegt bei $(1 - \beta) .08$ (siehe Tabelle C.29 im Anhang).

11.6 Weiterführende Untersuchungen

Zwei Ergebnisse der bisherigen Untersuchungen sind der Anlass dafür, dass an dieser Stelle noch weitere explorative Analysen durchgeführt werden, die nicht hypothesenüberprüfend sind und sich somit auch nicht direkt auf die Forschungsfragen beziehen. Der erste Grund sind die sehr gering gemessenen, impliziten und expliziten Assoziationen. Sie werfen die Frage auf, ob dies eine Folge der Mittelwertbildung über zu heterogene Gruppen ist, oder ob die Assoziationen tatsächlich so gering sind bzw. nicht sensibel genug gemessen wurden. Der zweite Grund sind die wenigen signifikanten Interaktionen der ANOVAs. Hierbei stellt sich die Frage, ob bei der Betrachtung anderer Zwischensubjektfaktoren mehr Aufklärung der Varianz erreicht werden kann.

Hinweise für die Beantwortung beider Fragen liefert eine exemplarische Untersuchung der drei IAT-Dimensionen mit dem Zwischensubjektfaktor *Lehrer*. In vielen vergleichbaren Studien zeichnet sich die Lehrervariable als besonders effektiv ab (Osbor-

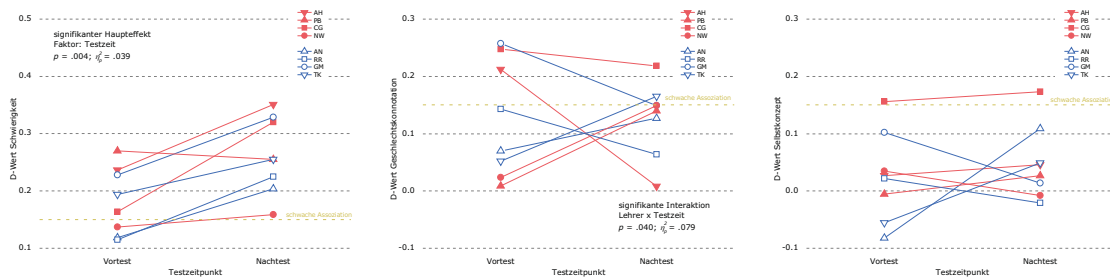


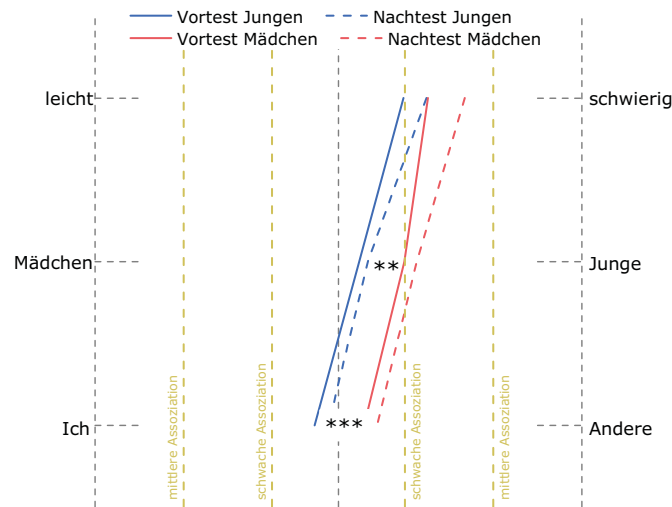
Abb. 11.9 Interaktionsdiagramme (Testzeit \times Lehrer) der drei IAT-Dimensionen *Schwierigkeit*, *Geschlechtskonnotation* und *Selbstkonzept* (v. li. n. re.).

ne, Simon & Collins, 2003). Somit liegt auch hier die Vermutung nahe, dass der Lehrer mehr Einfluss auf die Assoziationen zum Schulfach und deren Entwicklung hat, als das Curriculum.

Für alle *D*-Werte wurde eine zweifaktorielle Varianzanalyse mit Messwiederholung gerechnet. Der Innersubjektfaktor bleibt wie bei den bisherigen ANOVAs die Testzeit (zwei Stufen). Der Zwischensubjektfaktor *Lehrer* hat nun acht Stufen.

Bezüglich der ersten Frage, ob die geringen Assoziationen eine Folge der Mittelwertbildung sind, liefern alle Interaktionsdiagramme eine Antwort (Abb. 11.9). Die verhältnismäßig großen Spannweiten der Testergebnisse im Vor- oder Nachtest zeigen, dass die Varianz im hohen Maße durch den Unterricht bei einem bestimmten Lehrer aufgeklärt werden kann. Bei der Dimension *Schwierigkeit* zeigen sich beim Lehrer AH im Nachtest mittlere Assoziation ($M = 0.35$, $SE = 0.08$) und im Kontrast dazu bei NW nur schwache Assoziationen ($M = 0.16$, $SE = 0.06$). Beide Extrembeispiele sind aus der Versuchsgruppe (rot), wodurch bei der Mittelwertbildung die Assoziation maximal relativiert wird. Die Assoziation von *Physik* mit *schwierig* nimmt zwar bei fast allen Lehrern zu ($F(1, 211) = 8.55$, $p = .004$, $\eta_p^2 = .04$), die Stärke jedoch ist vom Lehrer abhängig. Ein ähnliches Bild der großen Spannweiten zeigt sich auch bei den anderen Dimensionen, obwohl die absoluten Ausprägungen der Assoziationen hier geringer sind. Dennoch zeigt sich, dass diese Ausprägung in Einzelfällen stark vom Mittelwert der restlichen Gruppe abweichen kann, wie z. B. bei Lehrer CG in der Dimension *Selbstkonzept* (Abb. 11.9 rechts).

Die zweite Frage betreffend, zeigt exemplarisch die Dimension *Geschlechtskonnotation*, inwieweit mehr Varianzaufklärung durch den Faktor *Lehrer* als durch den Faktor *Gruppe* erreicht wird. An dieser Stelle sei noch einmal angemerkt, dass alle Lehrpersonen männlich waren. Der Verlauf der Linien im Interaktionsdiagramm (Abb. 11.9 mittig) deutet darauf hin, dass sich die Assoziationen zu *Junge* zwar zwischen den Testzeitpunkten bei unterschiedlichen Lehrern auch unterschiedlich verändern, dies jedoch nicht von der Gruppenzugehörigkeit abhängig ist. Es kreuzen sich sowohl rote (Versuchsgruppe) als auch blaue Linien (Kontrollgruppe). Die zusätzliche Varianzaufklärung drückt sich in einer signifikanten Interaktion der beiden Faktoren aus, $F(7, 176) = 2.16$, $p = .040$, $\eta_p^2 = .08$ (mittlerer Effekt).



Weitere Anhaltspunkte für die Begründung der gering gemessenen Assoziationen liefern differenzierte Untersuchungen der Ergebnisse nach dem Geschlecht der Schülerinnen und Schüler. Für die Identifikation von Effekten, die von der Gruppenzugehörigkeit (VG und KG), vom Testzeitpunkt und zusätzlich vom Geschlecht abhängen, wurden für die drei IAT-Dimensionen dreifaktorielle Varianzanalysen mit Messwiederholung auf einem Faktor gerechnet. Diese entsprechen den ANOVAs aus Abschnitt 11.2 mit einem zusätzlichen Zwischensubjektfaktor *Geschlecht*.

Die Analysen zeigen mittlere Haupteffekte auf diesem Faktor für die Dimensionen *Geschlecht*, $F(1, 180) = 8.65$, $p = .004$, $\eta_p^2 = .05$ und *Selbstkonzept*, $F(1, 193) = 12.29$, $p = .001$, $\eta_p^2 = .06$ (siehe Abb. 11.10). Dies zeigt wieder, dass die gering gemessenen Assoziationen durch die Mittelwertbildung zu erklären sind, in diesem Fall über beide Geschlechter. Bei den Mädchen sind die stereotypischen Assoziationen stärker ausgeprägt, als bei den Jungen.

In der Dimension *Schwierigkeit* gibt es außerdem eine marginal signifikante Interaktion $\text{Testzeit} \times \text{Geschlecht} \times \text{Gruppe}$ mit schwacher Effektstärke, $F(1, 215) = 2.75$, $p = .099$, $\eta_p^2 = .01$. Dies ist ein Hinweis dafür, dass sich der Geschlechtsunterschied der Assoziation von *Physik* mit *schwierig* über die Zeit in Versuchs- und Kontrollgruppe unterschiedlich entwickelt (siehe Abb. 11.11). Während bei den Mädchen die Assoziation in beiden Gruppen steigt, nähert sie sich bei den Jungen in der Versuchsgruppe dem Niveau der Mädchen an (konvergiert). In der Kontrollgruppe bleibt sie konstant, womit dort die Assoziationen beider Geschlechter divergieren.

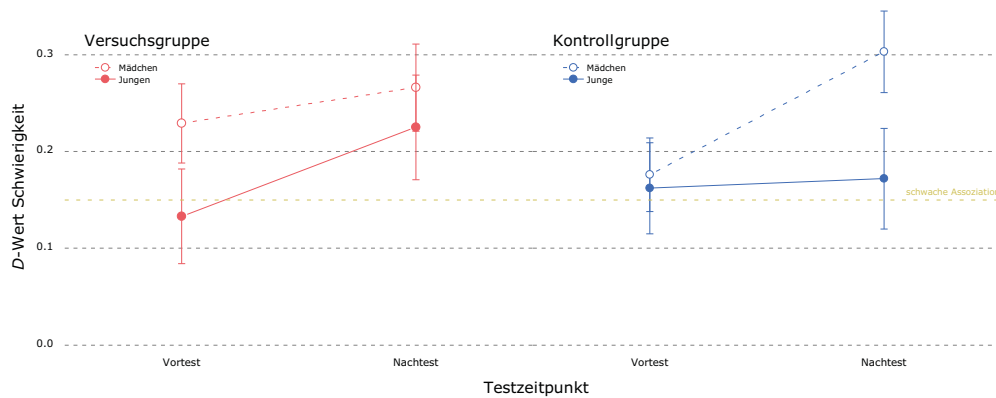


Abb. 11.11 Interaktionsdiagramm (Testzeit \times Geschlecht \times Gruppe) der Dimension *Schwierigkeit*

Teil III

Schluss

Diskussion und Ausblick

Der Einfluss phänomenbasierten Physikunterrichts auf das Lernen von Physik wurde in der vorliegenden Arbeit untersucht durch die Analyse der Auswirkungen einer phänomenbasierten Unterrichtssequenz der Anfangsoptik auf die impliziten und expliziten Assoziationen, auf das Interesse und auf das Fachwissen zur Optik. Im folgenden Abschnitt werden zunächst die Ergebnisse bezüglich der Assoziationen diskutiert und in Beziehung zu der bisherigen Forschung auf dem Gebiet gesetzt. Schließlich werden die entsprechenden Forschungsfragen 2 und 3 beantwortet und Grenzen der vorliegenden Studie aufgezeigt, welche die Messung der Assoziationen betreffen. Darauf folgend findet die Diskussion der Ergebnisse bezüglich des Interesses und des Fachwissens statt, einschließlich der Beantwortung von Forschungsfragen 1 und 4. In einem weiteren Abschnitt werden die Ergebnisse allgemein interpretiert und weitere Beschränkungen der vorliegenden Studie genannt. Aus den Ergebnissen, Schlussfolgerungen und Grenzen der vorliegenden Arbeit ergeben sich Ausblicke für weitere Untersuchungen.

12.1 Assoziationen zu Physik

Die mit den IATs gemessenen, impliziten Assoziationen zu *Physik* stützen die Ergebnisse von Kessels et al. (2006), Cvencek et al. (2011) und Nosek et al. (2002) in vergleichbaren Studien. Die positiven Vorzeichen der *D*-Werte zeigen, dass auch hier das Fach eher mit *schwierig*, *Junge* und *Andere* assoziiert wird als mit *leicht*, *Mädchen* und *Ich*. Qualitativ entsprechen die Assoziationen ebenfalls den Ergebnissen der Pilotierungen (Rehfeldt, 2012). Die Höhe der *D*-Werte entspricht jedoch insgesamt nur schwachen bis mittelstarken Assoziationen und liegt damit unter den Werten von Kessels et al. (2006) und denen der Pilotierung (Geschlechtskonnotation). Die *D*-Werte bei Grundschulern

(Geschlechtskonnotation und Selbstkonzept) sind mit denen der vorliegenden Studie vergleichbar (Cvencek et al., 2011). Da die hier untersuchte Stichprobe überwiegend aus Klassenstufe 7 stammt, liegt die Vermutung nahe, dass sich die stärker ausgeprägten stereotypischen Assoziationen erst im fortgeschrittenen Alter bilden, nachdem der Physikunterricht eine längere Zeit besucht wurde. Dafür sprechen auch die gemessenen zeitlichen Entwicklungen. Obwohl sich der Zuwachs zwischen Vor- und Nachtest nur bei der *Schwierigkeit* statistisch signifikant zeigt, ist in allen Dimensionen die Tendenz erkennbar (siehe Abb. 11.2 und 11.10). Diese Entwicklungen überdecken sogar die zu erwartenden Übungseffekte, aufgrund derer die *D*-Werte bei häufig durchgeführten IATs im Allgemeinen sinken (siehe Abschnitt 4.1.3).

Diese Übungseffekte könnten auch für die hier beobachtbare Reihenfolge der Assoziationsstärken verantwortlich sein (*Schwierigkeit* am stärksten, *Selbstkonzept* am schwächsten), da die IATs stets eine feste Reihenfolge hatten (siehe Abschnitt 10.2). Es kann hier jedoch nicht endgültig entschieden werden, ob dies der Fall ist oder ob die Assoziationen tatsächlich in dieser Stärke vorliegen. Die zunehmende Anzahl an aussortierten Teilnehmern im zweiten und dritten IAT spricht allerdings für diesen Einfluss auf die Validität des Tests.

Die explizit gemessenen Assoziationen bestätigen die impliziten insofern, als dass sie ebenfalls nicht sehr stark ausgeprägt sind. Die Qualität innerhalb der einzelnen Dimensionen unterscheidet sich jedoch, was sich auch in den nicht existierenden Korrelationen zwischen impliziter und expliziter Messung zeigt. In der vorliegenden Studie muss also davon ausgegangen werden, dass die IATs und die semantischen Differentiale unterschiedliche Konstrukte messen. Diese Unabhängigkeit ist nicht ungewöhnlich, da der Bereich der bisher beobachteten Korrelationen sehr groß ist ($-.25 \leq r \leq .60$, Hofmann et al., 2005). Die möglichen Gründe dafür sind ausführlich in Abschnitt 4.1.3 diskutiert und sie liegen sehr wahrscheinlich in den bekannten Moderationseffekten. Am plausibelsten ist der Umstand, dass die Assoziationen insgesamt nur schwach ausgeprägt sind, was nach Nosek et al. (2005) auch den Zusammenhang beider Maße schwächt. Eine weitere mögliche Moderatorvariable ist die Vertrautheit mit der Bewertung der Kategorien. Meiner Meinung nach ist es wahrscheinlich, dass die noch recht jungen Schülerinnen und Schüler sich bisher wenig selbstreflektierende Gedanken über die Assoziationen zu Physik gemacht haben, was die Korrelationen ebenfalls schwächen würde. Soziale Erwünschtheit wird als Ursache hier ausgeschlossen. Zum einen tritt sie auch in den bisherigen Meta-Analysen nicht als Moderator auf (Hofmann et al., 2005) und zum anderen wäre es für Schülerinnen und Schüler der 7. Klasse kein Tabu, Physik als schwierig oder männlich zu bewerten.

Forschungsfrage 2 zielte auf das Image von Physik als ein schwieriges und männlich konnotiertes Unterrichtsfach und den Einfluss von PbPU darauf. Laut Hypothese 2 schwächen die Assoziationen von *Physik* mit *schwierig* und *männlich* nach einer mehrwöchigen Intervention in Form von phänomenbasiertem Unterricht zur Anfangsoptik stärker ab als in der Kontrollgruppe. Dies wäre ebenfalls durch eine signifikante Interaktion *Testzeit* \times *Gruppe* zusammen mit der entsprechenden Veränderung der mittleren

D-Werte bzw. der Mittelwerte der semantischen Differentiale in beiden Gruppen gezeigt. Bei den beiden Image-Dimensionen treten diese Wechselwirkungen nicht auf. Die impliziten Assoziationen zu *schwierig* verstärken sich zwar im Laufe der Zeit, jedoch für beide Gruppen gleichermaßen. Die expliziten Assoziationen bleiben in beiden Gruppen vergleichbar konstant. Die Antwort auf Forschungsfrage 2 lautet also, dass auf Grundlage dieser Untersuchung phänomenbasierter Physikunterricht der Anfangsoptik keinen Einfluss auf das Image von Physik in Form von impliziten und expliziten Assoziationen zu *schwierig* und *männlich* hat. Hypothese 2 kann hier mit Ausnahme der Unterhypothese SH 2.a nicht bestätigt werden. Letztere wird durch die vorliegenden Daten insofern gestützt, als dass die typischen Assoziationen zum Image von Physik auch hier vorliegen.

Forschungsfrage 3 hatte das Anliegen, der Wirkung PbPUs auf das physikbezogene Selbstkonzept nachzugehen, welches als Assoziation von *Physik* mit *Ich* oder *Andere* interpretiert wurde. Nach Hypothese 3 wird die Assoziation zum *Selbst* als Folge einer mehrwöchigen Intervention in Form eines phänomenbasierten Unterrichts zur Anfangsoptik im Vergleich zu Kontrollgruppe gestärkt. Analog zu den Image-Dimensionen wäre dies durch die signifikante Interaktion *Testzeit* \times *Gruppe* und einer entsprechenden Veränderung der mittleren impliziten und expliziten Assoziationen in beiden Gruppen gezeigt. Die beobachtete Wechselwirkung bei den impliziten Assoziationen ist marginal signifikant. In der Kontrollgruppe verschlechtert sich das fachbezogene Selbstkonzept, während es in der Versuchsgruppe konstant bleibt. Die expliziten Assoziationen verschlechtern sich in beiden Gruppen gleichermaßen. Aufgrund der geringen Effektstärken und den nur schwach ausgeprägten Assoziationen lautet auf Grundlage dieser Untersuchung die Antwort auf Forschungsfrage 3, dass phänomenbasierter Physikunterricht der Anfangsoptik keinen Einfluss auf das fachbezogene Selbstkonzept hat, womit auch Hypothese 3 nicht bestätigt werden kann.

Zwischen Versuchs- und Kontrollgruppe sind im zeitlichen Verlauf der Assoziationen zu Physik keine Unterschiede messbar, die auf eine Wirkung des Treatments schließen lassen. Vielmehr zeigen die weiterführenden Untersuchungen, dass sich mehr systematische Varianz durch die Lehrervariable und durch Geschlechtsunterschiede aufklären lässt. Letztere stützen ebenfalls die Ergebnisse von Kessels et al. (2006), die den Grund dafür in der höheren Relevanz des Fachimages für das eigene Selbstkonzept bei den Mädchen sehen. Im Vergleich zu den Jungen ist das Wissen über das Image für sie wichtiger und somit bewusster, weil sie beim Abgleich mit ihrem Selbstkonzept „mehr zu verlieren“ haben. Eine stärkere aktuelle Präsenz des Wissens über das Image führt auch zu stärkeren Assoziationen diesbezüglich. Sowohl die auf den Lehrer als auch die auf das Geschlecht zurückführbaren Effekte zeigen jedoch, dass die eingesetzten IATs grundsätzlich sensitiv für die untersuchten Variablen sind und von äußeren Faktoren abhängige Veränderungen messen können.

An dieser Stelle sei daran erinnert, dass es sich bei den IATs um eine relative Messung der Assoziationen bezüglich einer zweiten Zielkategorie (Englisch) handelt. Daher

ist jede hier getroffene Aussage in der Weise relativierbar, als dass Veränderungen oder Unterschiede auch mit Assoziationen zum Englischunterricht erklärbar sind. Diese Einschränkung ist bei der Verwendung von IATs unumgänglich und unter der Annahme zu vernachlässigen, dass sich der Englischunterricht in Versuchs- und Kontrollgruppe nicht systematisch unterscheidet.

Eine weitere Beschränkung der Aussagekraft der IAT-Ergebnisse ergibt sich durch die in Abschnitt 10.3 geschilderten Probleme bei den Erhebungen. Die Durchführung der IATs an den Schulen erwies sich als nicht ideal und führte letztendlich auch zu einer hohen Quote an nicht verwendbaren Testergebnissen. Dies beeinflusst wiederum die Teststärken und somit die Wahrscheinlichkeit, einen existierenden Interaktionseffekt auch zu entdecken. Die abnehmenden *D*-Werte im zweiten und dritten IAT sind nicht nur durch Übungseffekte erklärbar, sondern auch dadurch, dass die Konzentration zunehmend nachlässt und die Reaktionszeit sich immer mehr zufällig verteilt. Vor allem die impliziten Assoziationen zum Selbstkonzept sind daher ebenfalls vor dem Hintergrund zu betrachten, dass sie eventuell aufgrund äußerer Bedingungen geschwächt sind. Der Vergleich mit den expliziten Assoziationen spricht allerdings nicht dafür.

12.2 Interesse und Fachwissen

Das Interesse an Physik und die Qualität der Motivation ist hier im Allgemeinen eher hoch als niedrig ausgeprägt. In allen Skalen ist jedoch ein negativer Trend beobachtbar, der bei der *Freude und Interesse an Naturwissenschaften*, der *Instrumentellen naturwissenschaftsbezogenen Motivation in Physik* und bei der *Qualität der Lernmotivation* auch statistisch signifikant ist. Mit diesen Ergebnissen reiht sich die vorliegende Arbeit ohne Widerspruch in die Befunde der IPN-Interessenstudie, PISA 2006 und dem IQB-Ländervergleich ein (Hoffmann et al., 1998; Jansen et al., 2013; OECD, 2007a). Obwohl hier das Geschlecht nicht differenziert betrachtet wurde, ist davon auszugehen, dass auch beim Interesse dieselben Unterschiede wie bei den Assoziationen vorliegen. Dies würde ebenfalls konform zu den Ergebnissen der IPN-Studie und des Ländervergleichs sein (siehe Abschnitt 3.1). Der Umstand, dass die Ausprägung des Interesses hier vergleichsweise hoch ist, wird durch die negative Tendenz zwischen Vor- und Nachtest relativiert. Somit ist davon auszugehen, dass dieser Trend stabil bleibt und auch die hier untersuchte Stichprobe am Ende der Sekundarstufe I auf dem Niveau von PISA 2006 liegt.

Die wahrgenommenen *Unterstützenden Lehr-/Lernbedingungen* und die *Qualität der Lernmotivation* entwickeln sich in der Versuchsgruppe nachteilig im Vergleich zur Kontrollgruppe. In beiden Fällen handelt es sich um schwache Effekte. Die erste Skala beinhaltet neun Aussagen über das Verhalten des Lehrers (z. B. „Die Lehrerin/ der Lehrer traut mir zu, dass ich die naturwissenschaftlichen Inhalte verstehen kann“, siehe A.2 im Anhang, Item 1 bis 9). Trotzdem klärt die Zugehörigkeit zu einer der beiden Gruppen mehr Varianz auf, als die Lehrervariable. Wenn man diesem (schwachen) Effekt die Bedeutsamkeit einer durch phänomenbasierten Physikunterricht systematisch hervorge-

rufenen Ursache zuordnet, folgt daraus, dass durch PbPU das Verhalten der Lehrperson beeinflusst wäre. Die Lehrperson würde demnach den psychologischen Grundbedürfnissen der Schülerinnen und Schüler nach sozialer Eingebundenheit, Kompetenzunterstützung und Autonomie weniger gerecht werden. Aufgrund der niedrigen Effektstärke ist es m. E. jedoch wahrscheinlicher, dass wegen der nicht zufälligen Zuordnung von Lehrern zur Versuchs- und Kontrollgruppe zwei oder mehr Lehrer innerhalb einer Gruppe extremes Ankreuzverhalten in dieser Skala verursachen.¹ Dies würde die Interaktion *Testzeit* \times *Gruppe* verstärken, auch wenn dies nicht unbedingt auf die Eigenschaften des Unterrichts, sondern auf die des Lehrers zurückzuführen ist.

Die zweite Skala und die darin enthaltene Subskala *Intrinsische/Interessierte Motiviertheit* enthalten Items mit Aussagen über den Unterricht („Im Unterricht in Physik macht mir der Unterricht Spaß“, siehe A.2 im Anhang, Item 28 bis 44), weshalb eher davon ausgegangen werden muss, dass sich das Antwortverhalten auf diesen zurückführen lässt. Die Kennwerte der Skalen entwickeln sich in der Versuchsgruppe ebenfalls negativ im Vergleich zur Kontrollgruppe. Eine theoretisch fundierte Post-Hoc-Erklärung für diese Beobachtung lässt sich nur schwer konstruieren, zumal der Effekt nur schwach und damit empirisch wenig bedeutsam ist. Die Aspekte der Subjektivität, Mediation und vor allem der Affektivität im PbPU würden eher dafür sprechen, dass die Inhalte als spannend wahrgenommen werden, die Schülerinnen und Schüler mit Freude und Spaß am Unterricht teilhaben und motiviert sind, sich weiter mit den Themen zu beschäftigen. Dennoch zeigen die Daten ein anderes Bild. Es ist meiner Meinung nach möglich, dass die normativ festgesetzten Merkmale PbPUs von den Schülerinnen und Schülern anders wahrgenommen werden, als von den Vertretern in Form von Fachdidaktikern und Lehrern (siehe unten). Eine Klärung dieses scheinbar widersprüchlichen Befunds kann nur durch weitere Untersuchungen geschehen. In diesen könnten die einzelnen Elemente PbPUs identifiziert werden, bei denen eine solche Diskrepanz auftritt. Im Rahmen dieser Arbeit bleibt diese Frage unbeantwortet.

Forschungsfrage 1 betraf die Wirkung PbPUs auf das Interesse der Schülerinnen und Schüler an Physik. Die zugehörige Hypothese 1 lautete, dass es sich positiv im Vergleich zur Kontrollgruppe nach einer mehrwöchigen Intervention in Form eines phänomenbasierten Unterrichts zur Anfangsoptik verändert. Dies würde sich durch eine signifikante Interaktion *Testzeit* \times *Gruppe* zusammen mit einer entsprechenden Veränderung der Gruppenmittelwerte zeigen. Bei keiner der erhobenen Interessesskalen ist die Entwicklung in der Versuchsgruppe besser als in der Kontrollgruppe. Gegenteilig ist es sogar so, dass dort, wo die notwendige Interaktion auftritt, die relative Entwicklung schlechter im Vergleich beider Gruppen ist. Mit den vorliegenden Daten kann Hypothese 1 somit nicht bestätigt werden. In Bezug auf Forschungsfrage 1 bedeutet dies, dass auf Grundlage dieser Untersuchung phänomenbasiertem Physikunterricht der Anfangsoptik kein positiver

¹Bei nur einem Lehrer als Ausreißer wäre die Interaktion *Testzeit* \times *Lehrer* signifikant, was für keine der Interessesskalen der Fall ist.

Einfluss auf das Interesse belegt werden kann. Im Falle der *Qualität der Lernmotivation* ist sogar davon auszugehen, dass PbPU einen negativen Einfluss hat.

Hinter Forschungsfrage 4 lag das Interesse der Kontrolle, ob die phänomenbasierte Unterrichtssequenz zur Anfangsoptik zu Nachteilen im Fachwissen führt, weil z. B. mehr Zeit des Unterrichts für Aspekte wie Exploration oder Affektivität aufgebracht wird oder bestimmte Inhalte wie das Modell des Lichtstrahls nicht thematisiert werden. Nach Hypothese 4 existieren diese Nachteile nicht und nach einer mehrwöchigen Intervention entwickelt sich das Fachwissen in Versuchs- und Kontrollgruppe gleichermaßen positiv. Dies wäre durch einen signifikanten Haupteffekt über die Testzeit sowie durch eine nicht existierende Interaktion *Testzeit* \times *Gruppe* gezeigt. Beide Unterhypothesen (SH 4.a und SH 4.b) werden durch die vorliegenden Daten insofern bestätigt, als dass das Fachwissen in beiden gleichermaßen Gruppen steigt. Die Antwort auf Forschungsfrage 4 lautet somit, dass phänomenbasierter Physikunterricht der Anfangsoptik zwar positiven Einfluss auf das Fachwissen hat, dieser Zuwachs jedoch im Vergleich zu einer Kontrollgruppe weder vorteilhaft noch nachteilig ist.

12.3 Allgemeine Interpretation

Die dargelegten Ergebnisse der Hauptstudie und die Antworten auf die Forschungsfragen zeigen, dass mit der vorliegenden Arbeit die hypothetisch vermutete Wirkung PbPUs auf das Interesse, das Image von Physik und das fachbezogene Selbstkonzept nicht nachgewiesen werden kann. Damit fehlt der Beleg einer Förderung dieser individuellen Voraussetzungen erfolgreichen Lernens (siehe Abschnitt 2.3). Es gibt jedoch auch bis auf eine Ausnahme keine Indizien dafür, dass PbPU erfolgreiches Lernen behindert. Somit bleibt die allgemeine Forschungsfrage (S. 89) insofern offen, als dass keine Auswirkungen auf das Lernen von Physik festgestellt werden können, die auf phänomenbasierten Unterricht zurückzuführen sind. Es bleibt also die Frage nach einer Begründung für den Widerspruch zwischen 1. langjährigen Erfahrungen und Eindrücken zur phänomenbasierten Anfangsoptik in der Schule bzw. Hochschullehre und 2. den empirischen Daten dieser Arbeit.

Da für keines der untersuchten Konstrukte Unterschiede feststellbar sind, ist es nicht sinnvoll, die Verwerfung jeder einzelnen Hypothese differenziert zu diskutieren. Die Herleitung der Hypothesen geschah theoriegeleitet und hierarchisch. Eine Verwerfung von Hypothese 2 würde theoretisch auch die Verwerfung der anderen Hypothesen begründen. Mit der Reduktion des allgemeinen Forschungsbefunds dieser Arbeit auf eine Diskussion bezüglich des Images von Physik wäre den hypothetisch angenommenen Zusammenhängen zum Selbstkonzept und Interesse zu viel Bedeutung beigemessen. Die Frage nach einer messbaren Wirksamkeit phänomenbasierten Physikunterrichts scheint grundlegender zu sein.

Die unter 1. erwähnten Erfahrungen und Eindrücke stammen im Wesentlichen aus der Perspektive von Lehrern, Didaktikern und Lehramtsstudierenden. Sie haben alle die

Gemeinsamkeit, dass sie den Unterricht aus einer Metaperspektive betrachten, welche die Vermittelbarkeit, die Sachstruktur und die didaktischen Überlegungen berücksichtigt. Außerdem besitzen sie im Allgemeinen überdurchschnittliche Kenntnisse auf dem Gebiet der Anfangsoptik, die es ihnen ermöglicht, auf dieser Metaebene den phänomenbasierten Physikunterricht zu bewerten, da der Vergleich mit alternativen Vorgehensweisen möglich ist. Die Rückmeldungen aus Lehrveranstaltungen und Lehrerfortbildungen zu dem Thema sowie die in Abschnitt 2.2 vorgestellte Expertenbefragung zeigen, dass auf dieser Ebene die Besonderheiten des phänomenbasierten Vorgehens deutlich erkennbar sind. Das betrifft nicht zwingend die Akzeptanz der Vor- oder Nachteile, bei denen durchaus Uneinigkeit bestehen kann. Die Unterschiede zu einem üblicherweise durchgeführten und wahrscheinlich von den meisten selbst erfahrenem Optikunterricht werden jedoch beim Blick durch die „didaktische Brille“ sichtbar.

Auf welche Art und Weise lässt sich diese Besonderheit auch aus Schülerperspektive feststellen? Die in der vorliegenden Arbeit getroffene Vermutung bestand darin, diese am ehesten im motivationalen und assoziativen Bereich zu finden. Es wurde ein relativ breites Spektrum an Skalen eingesetzt, die verschiedene Aspekte des Interesses und der Motivation berücksichtigen. Die Assoziationen wurden auf einer expliziten und einer impliziten, automatischen Ebene untersucht, womit auch die gedanklichen Verknüpfungen einbezogen sind, die unbewusst und unkontrolliert existieren. Doch selbst diese latente und subtile Wirkung von PbPU auf die Schülerinnen und Schüler war hier nicht messbar.

Es ist denkbar, dass aus Schülersicht nichts „Besonderes“ am phänomenbasierten Vorgehen zu erkennen ist. Zumindest die Erfahrungen mit Physikunterricht am Anfang der Sekundarstufe noch gering sind. Die eher fachinhaltlichen Aspekte *Modellfreiheit/Modellkompetenz* und *Exploration* könnten vielleicht erst dann würdigend wahrgenommen werden, wenn sie sich im Verlauf des Physikunterrichts bewähren. Der Wert eines dauerhaft gültigen Modellverständnisses, einer Beschreibung der Natur anstelle der Erklärung und eines Verstehens durch explorative, ordnende Erkenntnisprozesse wird demnach erst dann erkannt, wenn eine Reflexion des eigenen Naturverständnisses eintritt, also eher am Ende der Sekundarstufe II.

Genauso ist es möglich, dass die vom PbPU vorgenommenen Versuche, die Schülerinnen und Schüler und deren Sinneswahrnehmung stärker einzubeziehen (*Subjektivität*), zwischen der Lebenswelt und der Physikwelt zu vermitteln (*Mediation*) und den Physikunterricht zum emotionalen Erlebnis zu machen (*Affektivität*), entweder von den Schülerinnen und Schülern nicht wahrnehmbar oder von ihnen nicht gefordert ist. Letzteres wäre z. B. der Fall, wenn die Erwartungshaltung an den Physikunterricht von dem abweicht, was von PbPU geleistet wird. Die Hoffnung auf sachliches, objektives und reduktionistisches Vorgehen wäre im gewissen Maße enttäuscht. Meines Erachtens ist dieser Erklärungsansatz jedoch nicht fruchtbar. Zum einen widerspricht dies teilweise den Theorien zur Interessen- Motivation- und Selbstkonzeptförderung (Frenzel et al., 2009; Krapp, 1998; Möller & Trautwein, 2009; Schiefele, 2009). Zum anderen müsste

sich eine solche Enttäuschung auch durch negative Auswirkungen auf das Interesse oder Selbstkonzept zeigen.

Für mich ist es eher plausibel, dass die Merkmalsaspekte phänomenbasierten Physikunterrichts nach einem Modul zur Anfangsoptik nicht so herausstehend und schwerwiegend sind, dass sie die von vielen Faktoren abhängigen Konstrukte auch in dem Maße beeinflussen, dass es auf den Unterricht zurückführbar ist. Die emotionalen und affektiven Erlebnisse reichen anscheinend nicht aus, um messbare Effekte hervorzurufen. Auch die Vorteile der Subjektivität und der Vermittlung eröffnen sich für die Schülerinnen und Schüler vielleicht erst im weiteren Verlauf der Schulzeit, wenn der Physikunterricht und sein Nutzen immer mehr hinterfragt werden.

Die Daten zeigen in unabhängigen Messungen parallele Entwicklungen in beiden Gruppen. Die Anzahl an unterschiedlichen Instrumenten und unterschiedlichen Konstrukten zusammen mit den qualitativ gleichen Ergebnissen erhöht die Evidenz dafür, dass sich die Gruppen in keinem das Forschungsinteresse betreffenden Punkt unterscheiden. Ob sich dies nach längerer Zeit in anderen Klassenstufen ändert, ist spekulativ. Anhand der Ergebnisse der vorliegenden Studie muss davon ausgegangen werden, dass es eine Diskrepanz zwischen der Einschätzung von Lehrern bzw. Didaktikern und der Wahrnehmung PbPUs seitens der Schülerinnen und Schüler gibt. Das Motiv der vorliegenden Arbeit war es, einen empirischen Beleg für Wirkung phänomenbasierten Physikunterrichts auf das Lernen von Physik zu geben. Leider ist dies mit den gewonnenen Daten zunächst nicht möglich.

Die ausgebliebenen Unterschiedseffekte müssen auch hinsichtlich der Möglichkeiten des quasi-experimentellen Designs diskutiert werden, womit an dieser Stelle die Grenzen der Aussagekraft der vorliegenden Studie angesprochen werden sollen. Die Planung der Studie als Feldstudie mit natürlich vorgefundenen Klassenverbänden hatte den Vorteil, dass sich PbPU unter realen Bedingungen bewähren musste, was die externe Validität der Studie erhöht. Leider treten damit auch Störvariablen auf, die nicht kontrolliert werden können. Neben dem untersuchten Unterricht beeinflussen Klasseneffekte, die Lehrer und nicht zuletzt auch die äußeren Bedingungen der Lernumgebung die gemessenen Konstrukte sehr stark. Auch wenn durch relativ viele verschiedene Lehrer in beiden Gruppen versucht wurde, diese Einflüsse zu schwächen, so zeigt sich durch die weiterführenden Untersuchungen, dass die Varianzaufklärung durch die Lehrer sehr viel größer ist, als durch die Gruppenzugehörigkeit. Selbst wenn geringe Effekte aufgrund des phänomenbasierten Unterrichts auftreten würden, wären diese je nach Lehrer durch seinen größeren Einfluss verzerrt. In Anbetracht der möglichen Störvariablen gehen so die gesuchten Effekte „im Rauschen“ unter.

Die Wahl der relativ langen Intervention sollte sicherstellen, dass der Unterricht auch langfristig wirken kann, da, wie oben diskutiert, nicht davon auszugehen ist, dass einzelne Einheiten die Besonderheiten PbPUs ausreichend erkennbar machen. Die lange Dauer verursacht aber auch eine Konfundierung der Ereignisse. So ist hier nachträglich nicht feststellbar, ob nicht einzelne Schlüsselmomente der phänomenbasierten Anfangs-

optik doch die vermuteten Wirkungen hervorrufen und ob diese durch andere Elemente der Unterrichtssequenz wieder abgeschwächt werden.

Eine begrenzte Aussagekraft der vorliegenden Arbeit besteht auch darin, dass nur ein Beispiel phänomenbasierten Anfangsunterrichts betrachtet wurde. Die Gründe dafür sind in Abschnitt 2.5 erläutert. Trotzdem können die hier getroffenen Aussagen nicht ohne Weiteres auf andere Inhaltsgebiete der Physik und auf andere Klassenstufen verallgemeinert werden. Letzteres ist vor allem vor dem Hintergrund der oben diskutierten Entwicklung der Assoziationen zu Physik und des Interesses zu beachten, sowie der hypothetischen Entwicklung der Wahrnehmung PbPUs seitens der Schülerinnen und Schüler.

12.4 Ausblick

In der vorliegenden Arbeit wurde der Versuch unternommen, den aus verschiedenen Wurzeln und dadurch mit fehlender einheitlicher Theorie entstandenen phänomenbasierten Physikunterricht so zu charakterisieren, dass er den ursprünglichen Zielvorstellungen gerecht wird und gleichzeitig von anderen Unterrichtskonzeptionen abgrenzbar ist. Außerdem wurde versucht, PbPU aus dem Blickwinkel einer verbreiteten Lerntheorie zu betrachten und ihn mit seinen Eigenschaften in einem Modell des Lernens zu verorten. Damit ist der erste Schritt für eine nachträgliche lernpsychologische Fundierung getan. Bisher begründete sich das phänomenbasierte Vorgehen hauptsächlich aus erkenntnistheoretischen und fachinhaltlichen Überlegungen. Damit die Wirkungen auch empirisch überprüfbar sind, die den normativen Eigenschaften des Unterrichts unterstellt werden, ist es notwendig, eine begründete lerntheoretische Positionierung vorzunehmen, aus der heraus auch weitere phänomenbasierte Unterrichtskonzepte theoriegeleitet entstehen können. Eine solche einheitliche Theorie zum phänomenbasierten naturwissenschaftlichen Unterricht stellt ein wichtiges Forschungsdesiderat für zukünftige fachdidaktische Forschungen dar.

In Bezug auf die Wirksamkeit bestehender phänomenbasierter Unterrichtssequenzen bleibt zu erkunden, auf welcher Ebene die Schülerinnen und Schüler davon beim Lernen von Physik profitieren. Es ist also sinnvoll, hypothesengenerierende Forschungsvorhaben anzustreben, aus denen die Besonderheit PbPUs aus Schülerperspektive hervorgeht oder in denen beobachtbare Unterschiede exploriert werden. Dazu gehört auch die Identifikation einzelner Kriterien, aus denen Alleinstellungsmerkmale hervorgehen oder sogar ein Grad an Phänomenbasiertheit. Die Kriterien wären der erste Ansatz für weitere empirische Überprüfungen unter stärker kontrollierten Bedingungen.

Bei der Interessenskala *Qualität der Lernmotivation* gibt es ein Indiz dafür, dass PbPU einen negativen Einfluss hat. Auch dies wäre ein Angriffspunkt, um die Prozesse bei der Rezeption des Unterrichts seitens der Schülerinnen und Schüler besser zu verstehen. In weiteren unterrichtsbegleitenden Untersuchungen könnten die Ergebnisse dieser

Studie zunächst reproduziert werden, um sie anschließend z. B. mithilfe von Interviews besser einordnen zu können.

Die impliziten Assoziationstests wurden hier nach meinem Wissen erstmals in der naturwissenschaftsdidaktischen Forschung eingesetzt. Die große Flexibilität des IATs legt nahe, dass es weitere sinnvolle Möglichkeiten der Verwendung gibt, z. B. im Kontext der Balanced Identity Theory. Eine Adaption der Theorie bietet sich am ehesten auf dem Gebiet der pädagogischen Psychologie an, auf dem sie bisher unbeachtet scheint. Mit den Konstrukten des *Selbstkonzepts* und des *Selbstwerts* besteht jedoch eine große Schnittmenge, die auch in Form des fachbezogenen (Fähigkeits-)Selbstkonzepts für die fachdidaktische Forschung Relevanz hat. Damit besteht auch in Zukunft die Möglichkeit eines gewinnbringenden Einsatzes des IATs in vielen Fragestellungen. Dies ist vor allem vor dem Hintergrund zu beachten, dass in vielen Studien, und nicht zuletzt auch in der vorliegenden Arbeit, der Wert des IATs in Form von zusätzlicher Varianzaufklärung gezeigt wurde. Eine kategorische Beschränkung auf explizite Testverfahren wäre demnach fahrlässig.

Schließlich stellt sich aus fachdidaktischer Sicht auch die Frage, ob mit impliziten Assoziationen neben den Konstrukten *Einstellung*, *Stereotyp*, *Selbstwert* und *Selbstkonzept* auch weitere kognitive Verknüpfungen wie z. B. Schülervorstellungen oder die Vernetzung verschiedener Wissensbereiche erfasst werden können. Die assoziativen Wissensstrukturen (SKS-Diagramme) im Rahmen der BIT erinnern stark an Concept Maps, die häufig bei fachdidaktischen Fragestellungen eingesetzt werden. Es liegt also nahe, dass in diesen Fällen die Messungen sinnvoll mit IATs ergänzt werden können.

Die vorliegende Studie hat Probleme bei der Durchführung von IATs in Feldstudien aufgezeigt. Es bleibt also aus testadministrativer Sicht zu erforschen, unter welchen Bedingungen die IATs durchgeführt werden müssen, um hohe Ausschlussraten z. B. durch Konzentrationsmangel zu verhindern. Hier gilt es einen Kompromiss zu finden, zwischen unkompliziertem, flexiblem und schnellem Einsatz auf der einen Seite sowie zuverlässigen und validen Assoziationsmessungen auf der anderen Seite. Sollte sich der IAT als Instrument herausstellen, welches nur in Laborstudien eingesetzt werden kann, wäre sein Einsatzbereich in der fachdidaktischen Forschung stark eingeschränkt.

Inwieweit der IAT auch in zukünftigen Vorhaben gewinnbringend eingesetzt werden kann, hängt also von weiteren Untersuchungen zu seinem Einsatzbereich und der Testadministration ab. Die vorliegende Arbeit soll neben der Abhandlung der Forschungsfragen auch als Ausgangspunkt für diese beiden Richtungen angesehen werden.

Zusammenfassung

Die Motivation dieser Arbeit bestand darin, eine empirische Fundierung des Einflusses von phänomenbasierten Physikunterricht auf die Interessen- und Leistungsentwicklung bei Schülerinnen und Schülern vorzulegen. Dafür wurde zunächst der Untersuchungsgegenstand vorgestellt, indem die Merkmale phänomenbasierten Physikunterrichts aus den bisherigen Vorarbeiten gesammelt und übersichtlich dargestellt wurden. Da die bestehenden Konzeptionen und diesbezügliche Forderungen an Physikunterricht teilweise unterschiedliche Wurzeln, erkenntnistheoretische Positionierungen und Schwerpunktsetzungen haben, wurden für die Charakterisierung nur die Merkmalsaspekte herangezogen, die als Konsens so vieler Quellen wie möglich gelten. Das Ergebnis sind fünf Aspekte der Merkmale phänomenbasierten Physikunterrichts: 1. Subjektivität, 2. Mediation, 3. Affektivität, 4. Modellfreiheit/Modellkompetenz und 5. Exploration.

Anschließend wurde für ein weiteres theoriegeleitetes Vorgehen der PbPU aus dem Blickwinkel einer konstruktivistischen Lerntheorie betrachtet und es konnte aufgezeigt werden, dass beides miteinander vereinbar ist. Darauf aufbauend wurde ein Modell für erfolgreiches Lernen (INVO-Modell) herangezogen, in dem phänomenbasierter Physikunterricht mit seinen Merkmalen und Zielvorstellungen verortet werden konnte. Auf der Basis des Modells wurden die Ziele PbPUs gesammelt, denen das erfolgreiche Lernen als übergeordnetes Motiv dient. Ein Schwerpunkt stellt dabei die Förderung des Interesses an Physik dar.

Für die weiteren Betrachtungen wurde der Untersuchungsgegenstand auf die phänomenbasierte Anfangsoptik eingeengt. Das Curriculum wurde in seinem Ablauf grob dargestellt und seine Entstehung skizziert. Schließlich wurde es in Beziehung zum PbPU gesetzt und dargelegt, warum es sich für eine exemplarische Untersuchung eignet.

Nachdem das zu untersuchende Objekt vorgestellt war, wurden die zu berücksichtigten Konstrukte definiert und in Beziehung zueinander gesetzt. Das Interesse an Physik

tritt dabei als ein sehr gut erforschtes Konstrukt auf, weshalb der diesbezügliche Forschungsstand dargelegt wurde. Die Ergebnisse mehrerer Studien aber auch psychologische und bildungspolitische Gründe belegen die Relevanz der Förderung des Interesses. Im Vergleich zu anderen Fächern ist das Interesse an Physik gering, was besonders auf die geringe Ausprägung bei den Mädchen zurückführbar ist. An diesen Relationen und Geschlechtsunterschieden hat sich im Verlauf der letzten 20 Jahre wenig verändert.

Das Interesse an Physik wurde anschließend mit dem Image von Physik in Verbindung gebracht, welches als Prädiktor dafür angesehen werden kann. Daher wurde die Forschungslage zum Image von Physik erörtert, aus dem das Fach als *schwierig* und *männlich* hervorgeht. Beide Eigenschaften behindern vor allem bei Mädchen die Entstehung eines fachbezogenen Selbstkonzepts und Interesses. Außerdem wurde eine Beziehung zwischen dem Image und Assoziationen zu Physik hergestellt, die auf expliziter und impliziter Ebene für das Thema relevant sind.

Mit dem IAT und ST-IAT wurden zwei Instrumente eingeführt, die für die Messung der impliziten Assoziationen geeignet sind. Vor dem Hintergrund, dass es sich hierbei um Tests handelt, die in der fachdidaktischen Forschung unüblich sind, wurde sehr ausführlich der genaue Ablauf dargestellt und die Validität erörtert. Für die Messungen der impliziten Assoziationen wurde schließlich die Balanced Identity Theory (BIT) als Grundlage genommen, um aus ihr heraus Hypothesen für die Zusammenhänge zwischen dem Image von Physik und dem fachbezogenen Selbstkonzept zu generieren. Im Rahmen der Theorie werden die Konzepte als Assoziationen interpretiert, die mit den impliziten Tests messbar sind.

Aus den Merkmalen und Zielen phänomenbasierten Physikunterrichts, der lerntheoretischen Verortung, den wechselseitigen Zusammenhängen zwischen Interesse, Selbstkonzept und Assoziationen zu Physik sowie aus den Implikationen der BIT wurden die Forschungsfragen der vorliegenden Arbeit entwickelt, welche sich auf die Auswirkung phänomenbasierten Physikunterrichts auf das Lernen von Physik beziehen. Die dazugehörigen Hypothesen lauteten, dass sich PbPU positiv auf das Image von Physik, positiv auf das physikbezogene Selbstkonzept und positiv auf das Interesse an Physikunterricht auswirkt. In Bezug auf das Fachwissen wurde außerdem angenommen, dass PbPU keine nachteiligen Effekte verursacht.

In den folgenden Abschnitten wurde das Untersuchungsdesign, die Analysemethoden, die eingesetzten Instrumente und das Treatment vorgestellt. Die Untersuchung war als quasiexperimentelle Feldstudie mit Kontrollgruppendesign mit Vor- und Nachtest angelegt. Für die Auswertungen wurden verschiedene varianzanalytische Verfahren eingesetzt. In zwei Pilotstudien wurden die Instrumente und das Treatment geprüft. Da sich in beiden Pilotstudien der ST-IAT als ungeeignet und dem IAT unterlegen herausstellte, wurde fortan der IAT verwendet.

Die Ergebnisse der Hauptstudie zeigen bis auf eine Ausnahme bei den erhobenen Variablen keine Unterschiede zwischen den Gruppen, die auf den phänomenbasierten Physikunterricht zurückzuführen sind. Die absoluten Ausprägungen der gemessenen Konstrukte und die qualitativen Verläufe über die Zeit sind mit der Theorie und den bis-

herigen Forschungsergebnissen vereinbar. Lediglich die fehlenden Wechselwirkungen zwischen der Testzeit und der Gruppe führten dazu, dass die das Image, das Selbstkonzept und das Interesse betreffenden Hypothesen nicht bestätigt werden konnten. Im Wesentlichen muss davon ausgegangen werden, dass entweder PbPU bei den untersuchten Variablen keine Effekte hervorruft oder sie durch Störvariablen zu stark beeinflusst werden. Ersteres zieht die Frage nach sich, worin der Einfluss PbPUs bestehen könnte. Bezüglich der Alternative bleibt zu überlegen, wie die geeignete Methode der Feststellung aussehen muss. Weiterführende Untersuchungen konnten zeigen, dass sich das Geschlecht und die Lehrperson auf die Ausprägungen sowie die zeitliche Entwicklung der Variablen auswirkt. Beides sind Indizien dafür, dass die Instrumente grundsätzlich sensitiv für die untersuchten Variablen sind.

A.1 IATs und ST-IATs

Tabelle A.1 Übersicht der Stimuli in den drei eingesetzten IATs

Zieldimension Schulfach Physik	Attributdimensionen Schwierigkeit schwierig leicht	Geschlecht Junge Mädchen	Selbstkonzept Andere Ich
Formel	Vokabeln	kompliziert einfach	Jan Anna
Experiment	Grammatik	mühevoll simpel	Lukas Julia
Mechanik	Fremdsprache	mühsam locker	Tim Laura
Masse	Aussprache	anstrengend verständlich	Jonas Lena
Elektrizität	Rechtschreibung	schwer kinderleicht	Alexander Sara
			Ihr Euch
			Dein Ihr
			Mir Meine
			Selbst Mich
			Eigen

Inquisit-Skript für den IAT

Folgend ist der Quellcode für den *Schwierigkeit*-IAT aufgeführt. Hier wurden Ziel- und Attributdimensionen vertauscht, damit die Attributkategorien die Seite wechseln. Für die anderen Dimensionen müssen nur die Zeilen 29-49 und die *instructions* angepasst werden. Falls für einen zweiten und dritten IAT Übungsblöcke gestrichen werden sollen (siehe Abschnitt 7.2), erfolgt dies in den Zeilen 386ff.

```

1 <item attributeAlabel>
2 / 1 = "Physik"
3 </item>
4
5 <item attributeA>
6 / 1 = "Formel"
7 / 2 = "Experiment"
8 / 3 = "Mechanik"
```

```
9 / 4 = "Masse"
10 / 5 = "Elektrizitaet"
11 </item>
12
13 <item attributeBlabel>
14 /1 = "Englisch"
15 </item>
16
17 <item attributeB>
18 /1 = "Vokabeln"
19 /2 = "Grammatik"
20 /3 = "Fremdsprache"
21 /4 = "Aussprache"
22 /5 = "Rechtschreibung"
23 </item>
24
25 <item targetAlabel>
26 /1 = "schwierig"
27 </item>
28
29 <item targetA>
30 /1 = "kompliziert"
31 /2 = "muehevoll"
32 /3 = "muehsam"
33 /4 = "anstrengend"
34 /5 = "schwer"
35 </item>
36
37 <item targetBlabel>
38 /1 = "leicht"
39 </item>
40
41 <item targetB>
42 /1 = "einfach"
43 /2 = "simpel"
44 /3 = "locker"
45 /4 = "verstaendlich"
46 /5 = "kinderleicht"
47 </item>
48
49
50 *****
51 Task instruction stimuli
52 *****
53
54 <trial instructions>
55 / stimulustimes = [1=instructions , spacebar]
56 / correctresponse = (" ")
57 / errormessage = false
58 / recorddata = false
59 </trial>
60
61 <text instructions>
62 / items = instructions
63 / hjustify = left
64 / size = (90%, 60%)
65 / position = (50%, 85%)
66 / valign = bottom
67 / select = sequence
68 / resetinterval = 20
69 </text>
70
71 <item instructions>
```


- 72 / 1 = "Lege deine beiden Mittel- oder Zeigefinger auf die W- und P-Taste deiner Tastatur. In der Mitte des Bildschirms erscheinen nacheinander Woerter, die zu den Bereichen 'schwierig' und 'leicht' gehoeren. Wenn das Wort zu dem Bereich auf der linken Seite gehoert, druecke das 'W', wenn es zu dem Bereich auf der rechten Seite gehoert, druecke das 'P'. Ein Wort gehoert immer eindeutig zu nur einem Bereich. Wenn Du einen Fehler machst, erscheint ein X – druecke dann die andere Taste.
- 73
- 74 Bei dieser Aufgabe wird die Zeit gemessen. Druecke immer so SCHNELL WIE MoeGLICH und mache so WENIGE FEHLER WIE MoeGLICH. Der Test dauert etwa 5 Minuten."
- 75
- 76 / 2 = "Wie du oben siehst, habe sich die Bereiche nun geaendert. Die Worte, die in der Mitte erscheinen, werden ebenfalls andere sein, die Regeln bleiben aber dieselben.
- 77
- 78 Wenn das Wort zu dem Bereich auf der linken Seite gehoert, druecke das 'W', wenn es zu dem Bereich auf der rechten Seite gehoert, druecke das 'P'. Ein Wort gehoert immer eindeutig zu nur einem Bereich. Wenn Du einen Fehler machst, erscheint ein X – druecke dann die andere Taste. Druecke immer so SCHNELL WIE MoeGLICH und mache so WENIGE FEHLER WIE MoeGLICH."
- 79
- 80 / 3 = "Wie du oben siehst, erscheinen nun alle vier Bereiche zusammen. Beachte aber, dass jedes Wort eindeutig zu nur einem Bereich gehoert! Wenn z.B. 'Physik' und 'schwierig' oben auf zwei verschiedenen Seiten erscheinen, muessten Begriffe, die zu 'Physik' gehoeren (z.B. 'Masse') der Seite, wo 'Physik' steht, zugeordnet werden.
- 81
- 82 Die farbigen Markierungen der Begriffe (gruen und weiss) helfen Dir bei der Zuordnung. Benutze die Tasten 'W' und 'P' fuer die Sortierung zu den Bereichen nach links oder rechts. Berichtige deine Fehler durch Druecken der anderen Taste."
- 83
- 84 / 4 = "Sortiere dieselben vier Bereiche noch einmal. Bitte sei so SCHNELL WIE MoeGLICH MIT SO WENIG FEHLERN WIE MoeGLICH.
- 85
- 86 Die farbigen Markierungen der Begriffe (gruen und weiss) helfen Dir bei der Zuordnung. Benutze wieder die Tasten 'W' und 'P' fuer die Sortierung zu den Bereichen nach links oder rechts. Berichtige wieder deine Fehler durch Druecken der anderen Taste."
- 87
- 88 / 5 = "Achtung, jetzt erscheinen wieder nur zwei Bereiche und die haben ihren Platz getauscht. Der Bereich, der vorher links war, ist nun rechts und der Bereich der vorher rechts war, ist nun links. uebe nun die neue Anordnung.
- 89
- 90 Benutze die Tasten 'W' und 'P' fuer die Sortierung zu den Bereichen nach links oder rechts. Berichtige deine Fehler durch Druecken der anderen Taste."
- 91
- 92 / 6 = "Wie du oben siehst, erscheinen nun alle vier Bereiche in der neuen Anordnung. Wieder gehoert jedes Wort eindeutig zu nur einem Bereich.
- 93
- 94 Die farbigen Markierungen der Begriffe (gruen und weiss) helfen Dir bei der Zuordnung. Benutze die Tasten 'W' und 'P' fuer die Sortierung zu den Bereichen nach links oder rechts. Berichtige deine Fehler durch Druecken der anderen Taste."
- 95
- 96 / 7 = "Sortiere dieselben vier Bereiche noch einmal. Bitte sei so SCHNELL WIE MoeGLICH MIT SO WENIG FEHLERN WIE MoeGLICH.
- 97
- 98 Die farbigen Markierungen der Begriffe (gruen und weiss) helfen Dir bei der Zuordnung. Benutze wieder die Tasten 'W' und 'P' fuer die Sortierung zu den Bereichen nach links oder rechts. Berichtige wieder deine Fehler durch Druecken der anderen Taste."
- 99

```
100 </item>
101
102 <text spacebar>
103 / items = ("Zum Starten die Leertaste druecken!")
104 / position = (50%, 95%)
105 / valign = bottom
106 </text>
107
108 <text attributeA>
109 / items = attributeA
110 / fontstyle = ("Arial", 5%)
111 / txcolor = (0, 255, 0)
112 </text>
113
114 <text attributeB>
115 / items = attributeB
116 / fontstyle = ("Arial", 5%)
117 / txcolor = (0, 255, 0)
118 </text>
119
120 <text targetB>
121 / items = targetB
122 / fontstyle = ("Arial", 5%)
123 </text>
124
125 <text targetA>
126 / items = targetA
127 / fontstyle = ("Arial", 5%)
128 </text>
129
130 <text error>
131 / position = (50%, 75%)
132 / items = ("X")
133 / color = (255, 0, 0)
134 / fontstyle = ("Arial", 10%, true)
135 </text>
136
137 <text attributeAleft>
138 / items = attributeAlabel
139 / valign = top
140 / halign = left
141 / position = (5%, 5%)
142 / txcolor = (0, 255, 0)
143 / fontstyle = ("Arial", 5%)
144 </text>
145
146 <text attributeBright>
147 / items = attributeBlabel
148 / valign = top
149 / halign = right
150 / position = (95%, 5%)
151 / txcolor = (0, 255, 0)
152 / fontstyle = ("Arial", 5%)
153 </text>
154
155 <text attributeAleftmixed>
156 / items = attributeAlabel
157 / valign = top
158 / halign = left
159 / position = (5%, 19%)
160 / txcolor = (0, 255, 0)
161 / fontstyle = ("Arial", 5%)
162 </text>
```

```
163
164 <text attributeBrightmixed>
165 / items = attributeBlabel
166 / valign = top
167 / halign = right
168 / position = (95%, 19%)
169 / txcolor = (0, 255, 0)
170 / fontstyle = ("Arial", 5%)
171 </text>
172
173 <text targetBleft>
174 / items = targetBlabel
175 / valign = top
176 / halign = left
177 / position = (5%, 5%)
178 / fontstyle = ("Arial", 5%)
179 </text>
180
181 <text targetBright>
182 / items = targetBlabel
183 / valign = top
184 / halign = right
185 / position = (95%, 5%)
186 / fontstyle = ("Arial", 5%)
187 </text>
188
189 <text targetAleft>
190 / items = targetAlabel
191 / valign = top
192 / halign = left
193 / position = (5%, 5%)
194 / fontstyle = ("Arial", 5%)
195 </text>
196
197 <text targetArigh>
198 / items = targetAlabel
199 / valign = top
200 / halign = right
201 / position = (95%, 5%)
202 / fontstyle = ("Arial", 5%)
203 </text>
204
205 <text orleft>
206 / items = ("oder")
207 / valign = top
208 / halign = left
209 / position = (5%, 12%)
210 / fontstyle = ("Arial", 5%)
211 </text>
212
213 <text orright>
214 / items = ("oder")
215 / valign = top
216 / halign = right
217 / position = (95%, 12%)
218 / fontstyle = ("Arial", 5%)
219 </text>
220
221 *****
222 Trials
223 *****
224
225 <trial attributeA>
```

```
226 / validresponse = ("W", "P")
227 / correctresponse = ("W")
228 / stimulusframes = [1 = attributeA]
229 / posttrialpause = 250
230 </trial>
231
232 <trial attributeB>
233 / validresponse = ("W", "P")
234 / correctresponse = ("P")
235 / stimulusframes = [1 = attributeB]
236 / posttrialpause = 250
237 </trial>
238
239
240 <trial targetBleft>
241 / validresponse = ("W", "P")
242 / correctresponse = ("W")
243 / stimulusframes = [1 = targetB]
244 / posttrialpause = 250
245 </trial>
246
247 <trial targetBright>
248 / validresponse = ("W", "P")
249 / correctresponse = ("P")
250 / stimulusframes = [1 = targetB]
251 / posttrialpause = 250
252 </trial>
253
254 <trial targetAleft>
255 / validresponse = ("W", "P")
256 / correctresponse = ("W")
257 / stimulusframes = [1 = targetA]
258 / posttrialpause = 250
259 </trial>
260
261 <trial targetAright>
262 / validresponse = ("W", "P")
263 / correctresponse = ("P")
264 / stimulusframes = [1 = targetA]
265 / posttrialpause = 250
266 </trial>
267
268 *****
269 Blocks
270 *****
271
272 <block attributepractice>
273 / bgstim = (attributeAleft, attributeBright)
274 / trials = [1=instructions;2-21 = random(attributeA, attributeB)]
275 / errormessage = true(error,200)
276 / responsemode = correct
277 </block>
278
279 <block targetcompatiblepractice>
280 / bgstim = (targetAleft, targetBright)
281 / trials = [1=instructions;2-21 = random(targetAleft, targetBright)]
282 / errormessage = true(error,200)
283 / responsemode = correct
284 </block>
285
286 <block targetincompatiblepractice>
287 / bgstim = (targetAright, targetBleft)
288 / trials = [1=instructions;2-21 = random(targetAright, targetBleft)]
```

```

289 / errormessage = true(error,200)
290 / responsemode = correct
291 </block>
292
293 <block targetcompatiblepracticeswitch>
294 / bgstim = (targetAleft , targetBright)
295 / trials = [1=instructions;2-41 = random(targetAleft , targetBright)]
296 / errormessage = true(error,200)
297 / responsemode = correct
298 </block>
299
300 <block targetincompatiblepracticeswitch>
301 / bgstim = (targetAright , targetBleft)
302 / trials = [1=instructions;2-41 = random(targetAright , targetBleft)]
303 / errormessage = true(error,200)
304 / responsemode = correct
305 </block>
306
307 <block compatibletest1>
308 / bgstim = (targetAleft , orleft , attributeAleftmixed , targetBright , orright ,
    attributeBrightmixed)
309 / trials = [1=instructions;
310   3,5,7,9,11,13,15,17,19,21= random(targetAleft , targetBright);
311   2,4,6,8,10,12,14,16,18,20 = random(attributeA , attributeB)]
312 / errormessage = true(error,200)
313 / responsemode = correct
314 / ontrialend = [if (block.compatibletest1.latency <= 10000 && block.compatibletest1
    .currenttrialnumber != 1 ) values.sum1a = values.sum1a + block.compatibletest1.
    latency]
315 / ontrialend = [if (block.compatibletest1.latency <= 10000 && block.compatibletest1
    .currenttrialnumber != 1 ) values.n1a = values.n1a + 1]
316 / ontrialend = [if (block.compatibletest1.latency <= 10000 && block.compatibletest1
    .currenttrialnumber != 1 ) values.s1a = values.s1a + (block.compatibletest1.
    latency * block.compatibletest1.latency)]
317 </block>
318
319 <block compatibletest2>
320 / bgstim = (targetAleft , orleft , attributeAleftmixed , targetBright , orright ,
    attributeBrightmixed)
321 / trials = [
322   2,4,6,8,10,12,14,16,18,20,22,24,26,28,30,32,34,36,38,40 = random(targetAleft ,
    targetBright);
323   1,3,5,7,9,11,13,15,17,19,21,23,25,27,29,31,33,35,37,39 = random(attributeA ,
    attributeB)]
324 / errormessage = true(error,200)
325 / responsemode = correct
326 / ontrialend = [if (block.compatibletest2.latency <= 10000) values.sum1b = values.
    sum1b + block.compatibletest2.latency]
327 / ontrialend = [if (block.compatibletest2.latency <= 10000) values.n1b = values.
    n1b + 1]
328 / ontrialend = [if (block.compatibletest2.latency <= 10000) values.s1b = values.
    s1b + (block.compatibletest2.latency * block.compatibletest2.latency)]
329 </block>
330
331
332 <block incompatibletest1>
333 / bgstim = (targetBleft , orleft , attributeAleftmixed , targetAright , orright ,
    attributeBrightmixed)
334 / trials = [1=instructions;
335   3,5,7,9,11,13,15,17,19,21 = random(targetBleft , targetAright);
336   2,4,6,8,10,12,14,16,18,20 = random(attributeA , attributeB)]
337 / errormessage = true(error,200)
338 / responsemode = correct

```

```
339 / ontrialend = [if (block.incompatibletest1.latency <= 10000 && block.
    incompatibletest1.currenttrialnumber != 1) values.sum2a = values.sum2a + block.
    incompatibletest1.latency]
340 / ontrialend = [if (block.incompatibletest1.latency <= 10000 && block.
    incompatibletest1.currenttrialnumber != 1) values.n2a = values.n2a + 1]
341 / ontrialend = [if (block.incompatibletest1.latency <= 10000 && block.
    incompatibletest1.currenttrialnumber != 1) values.ss2a = values.ss2a + (block.
    incompatibletest1.latency * block.incompatibletest1.latency)]
342 </block>
343
344 <block incompatibletest2>
345 / bgstim = (targetBleft, orleft, attributeAleftmixed, targetAright, orright,
    attributeBrightmixed)
346 / trials = [
347     2,4,6,8,10,12,14,16,18,20,22,24,26,28,30,32,34,36,38,40 = random(targetBleft,
        targetAright);
348     1,3,5,7,9,11,13,15,17,19,21,23,25,27,29,31,33,35,37,39 = random(attributeA,
        attributeB)]
349 / errormessage = true(error,200)
350 / responsemode = correct
351 / ontrialend = [if (block.incompatibletest2.latency <= 10000) values.sum2b =
    values.sum2b + block.incompatibletest2.latency]
352 / ontrialend = [if (block.incompatibletest2.latency <= 10000) values.n2b = values.
    n2b + 1]
353 / ontrialend = [if (block.incompatibletest2.latency <= 10000) values.ss2b = values
    .ss2b + (block.incompatibletest2.latency * block.incompatibletest2.latency)]
354 </block>
355
356 <block compatibletestinstructions>
357 / bgstim = (targetAleft, orleft, attributeAleftmixed, targetBright, orright,
    attributeBrightmixed)
358 / trials = [1=instructions]
359 / recorddata = false
360 </block>
361
362 <block incompatibletestinstructions>
363 / bgstim = (targetBleft, orleft, attributeAleftmixed, targetAright, orright,
    attributeBrightmixed)
364 / trials = [1=instructions]
365 / recorddata = false
366 </block>
367
368 <block summary>
369 / trials = [1=summary]
370 / recorddata = false
371 </block>
372
373 *****
374 Experiment
375 *****
376
377 <defaults>
378 / fontstyle = ("Arial", 3.5%)
379 / screencolor = (0,0,0)
380 / txbgcolor = (0,0,0)
381 / txcolor = (255, 255, 255)
382 / minimumversion = "3.0.0.0"
383 </defaults>
384
385 <expt>
386 / blocks = [1=block1; 2=attributepractice; 3=block3; 4=block4; 5=block5; 6=block6;
    7=block7; 8=block8; 9=block9]
387 </expt>
```

```

388
389 <variables>
390 / group = (1 of 2) (block1=targetcompatiblepractice; block3=compatibletest1; block4=
    compatibletestinstructions; block5=compatibletest2; block6=
    targetincompatiblepractice; block7=incompatibletest1; block8=
    incompatibletestinstructions; block9=incompatibletest2)
391 / group = (2 of 2) (block1=targetincompatiblepractice; block3=incompatibletest1;
    block4=incompatibletestinstructions; block5=incompatibletest2; block6=
    targetcompatiblepractice; block7=compatibletest1; block8=
    compatibletestinstructions; block9=compatibletest2)
392 </variables>
393
394 *****
395 Data Columns
396 *****
397
398 <data>
399 / columns = [date, time, subject, blockcode, blocknum, trialcode, trialnum, response
    , correct, latency, stimulusnumber, stimulusitem, expressions.da, expressions.db
    , expressions.d ]
400 </data>
401
402 *****
403 Test Monkey
404 *****
405 <monkey>
406 / latencydistribution = normal(500, 100)
407 / percentcorrect = 90
408 </monkey>
409
410 *****
411 Scoring
412 *****
413 <values>
414 / sum1a = 0
415 / sum2a = 0
416 / sum1b = 0
417 / sum2b = 0
418 / n1a = 0
419 / n2a = 0
420 / n1b = 0
421 / n2b = 0
422 / ss1a = 0
423 / ss2a = 0
424 / ss1b = 0
425 / ss2b = 0
426 / magnitude = "unknown"
427 </values>
428
429 * 1 is compatible, 2 is incompatible
430 * a is first block, b is second block
431
432 <expressions>
433 / m1a = values.sum1a / values.n1a
434 / m2a = values.sum2a / values.n2a
435 / m1b = values.sum1b / values.n1b
436 / m2b = values.sum2b / values.n2b
437 / sd1a = sqrt((values.ss1a - (values.n1a * (expressions.m1a * expressions.m1a))) / (
    values.n1a - 1))
438 / sd2a = sqrt((values.ss2a - (values.n2a * (expressions.m2a * expressions.m2a))) / (
    values.n2a - 1))
439 / sd1b = sqrt((values.ss1b - (values.n1b * (expressions.m1b * expressions.m1b))) / (
    values.n1b - 1))

```

```

440 / sd2b = sqrt(((values.ss2b - (values.n2b * (expressions.m2b * expressions.m2b))) / (
      values.n2b - 1))
441 / sda = sqrt((((values.n1a - 1) * (expressions.sd1a * expressions.sd1a) + (values.
      n2a - 1) * (expressions.sd2a * expressions.sd2a)) + ((values.n1a + values.n2a) *
      ((expressions.m1a - expressions.m2a) * (expressions.m1a - expressions.m2a)) /
      4) ) / (values.n1a + values.n2a - 1) )
442 / sdb = sqrt((((values.n1b - 1) * (expressions.sd1b * expressions.sd1b) + (values.
      n2b - 1) * (expressions.sd2b * expressions.sd2b)) + ((values.n1b + values.n2b) *
      ((expressions.m1b - expressions.m2b) * (expressions.m1b - expressions.m2b)) /
      4) ) / (values.n1b + values.n2b - 1) )
443 / da = (m2a - m1a) / expressions.sda
444 / db = (m2b - m1b) / expressions.sdb
445 / d = (expressions.da + expressions.db) / 2
446 / preferred = "unknown"
447 / notpreferred = "unknown"
448 </expressions>

```

Inquisit-Skript für den ST-IAT

Folgend ist der Quellcode für den *Schwierigkeit*-ST-IAT aufgeführt. Für die anderen Dimensionen müssen nur die Zeilen 2-22 und die *instructions* angepasst werden.

```

1 <item attributeAlabel>
2 /1 = "schwierig"
3 </item>
4
5 <item attributeA>
6 /1 = "kompliziert"
7 /2 = "muehevoll"
8 /3 = "unverstaendlich"
9 /4 = "anstrengend"
10 /5 = "schwer"
11 </item>
12
13 <item attributeBlabel>
14 /1 = "leicht"
15 </item>
16
17 <item attributeB>
18 /1 = "einfach"
19 /2 = "simpel"
20 /3 = "locker"
21 /4 = "muehelos"
22 /5 = "anspruchlos"
23 </item>
24
25 <item targetAlabel>
26 /1 = "Physik"
27 </item>
28
29 <item targetA>
30 / 1 = "Formel"
31 / 2 = "Experiment"
32 / 3 = "Mechanik"
33 / 4 = "Masse"
34 / 5 = "Elektrizitaet"
35 </item>
36
37
38 <trial summary>
39 / stimulustimes = [0=summary]

```



```

40 / validresponse = ( " " )
41 / recorddata = false
42 / ontrialbegin = [values.magnitude = "little to no"]
43 / ontrialbegin = [if( abs(expressions.d) > 0.15 ) values.magnitude = "geringfuegigen
    "]
44 / ontrialbegin = [if( abs(expressions.d) > 0.35 ) values.magnitude = "mittleren"]
45 / ontrialbegin = [if( abs(expressions.d) >= 0.65 ) values.magnitude = "starken"]
46 / ontrialbegin = [if(expressions.d >= 0.0) expressions.attitude = "schwierig"]
47 / ontrialbegin = [if(expressions.d < 0.0) expressions.attitude = "leicht"]
48 </trial>
49
50 <text summary>
51 / items = ("Ihr IAT-Wert (D) betraegt <% expressions.d %>, was einer <% values.
    magnitude %> impliziten Assoziation von <% expressions.attitude %> mit <% item.
    targetALabel.1 %> entspricht.~n~n~nDruecken Sie zum Beenden die Leertaste.")
52 / size = (60%, 60%)
53 / hjustify = left
54 </text>
55
56
57 <trial instructions>
58 / stimulustimes = [1=instructions , spacebar]
59 / correctresponse = ( " " )
60 / errormessage = false
61 / recorddata = false
62 </trial>
63
64 <text instructions>
65 / items = instructions
66 / hjustify = left
67 / size = (90%, 60%)
68 / position = (50%, 85%)
69 / valign = bottom
70 / select = sequence
71 / resetinterval = 20
72 </text>
73
74 <item instructions>
75 / 1 = "Lege deine beiden Mittel- oder Zeigefinger auf die W- und P-Taste deiner
    Tastatur. In der Mitte des Bildschirms erscheinen nacheinander Woerter, die zu
    den Bereichen 'schwierig' und 'leicht' gehoeren. Wenn das Wort zu dem Bereich
    auf der linken Seite gehoert, druecke das 'W', wenn es zu dem Bereich auf der
    rechten Seite gehoert, druecke das 'P'. Ein Wort gehoert immer eindeutig zu nur
    einem Bereich. Wenn Du einen Fehler machst, erscheint ein X - druecke dann die
    andere Taste.
76
77 Bei dieser Aufgabe wird die Zeit gemessen. Druecke immer so SCHNELL WIE MoeGLICH und
    mache so WENIG FEHLER WIE MoeGLICH. Der Test dauert etwa 5 Minuten."
78
79 / 2 = "Wie du oben siehst, erscheint nun ein neuer Bereich 'Physik' zusammen mit
    einem der beiden vorherigen. Wie eben schon gehoert jedes Wort eindeutig zu nur
    einem Bereich. Wenn z.B. 'Physik' und 'schwierig' oben auf zwei verschiedenen
    Seiten erscheinen, muessten Begriffe, die zu 'Physik' gehoeren (z.B. 'Masse')
    der Seite, wo 'Physik' steht zugeordnet werden.
80
81 Die farbigen Markierungen der Begriffe (gruen und weiss) helfen Dir bei der
    Zuordnung. Benutze wieder die Tasten 'W' und 'P' fuer die Sortierung zu den
    Bereichen nach links oder rechts. Berichtige wieder deine Fehler durch Druecken
    der anderen Taste."
82
83 / 3 = "Sortiere dieselben drei Bereiche noch einmal. Bitte sei so SCHNELL WIE
    MoeGLICH MIT SO WENIG FEHLERN WIE MoeGLICH.
84

```

```
85 Die farbigen Markierungen der Begriffe (gruen und weiss) helfen Dir bei der
    Zuordnung. Benutze wieder die Tasten 'W' und 'P' fuer die Sortierung zu den
    Bereichen nach links oder rechts. Berichtige wieder deine Fehler durch Druecken
    der anderen Taste."
86
87 / 4 = "Wie du oben siehst, stehen dieselben drei Bereiche in einer neuen
    Zusammensetzung. Denke daran: Jeder Begriff gehoert nur zu einer Gruppe.
88
89 Die farbigen Markierungen der Begriffe (gruen und weiss) helfen Dir bei der
    Zuordnung. Benutze wieder die Tasten 'W' und 'P' fuer die Sortierung zu den
    Bereichen nach links oder rechts. Berichtige wieder deine Fehler durch Druecken
    der anderen Taste."
90
91 / 5 = "Sortiere dieselben drei Bereiche noch einmal. Bitte sei so SCHNELL WIE
    MoeGLICH MIT SO WENIG FEHLERN WIE MoeGLICH.
92
93 Die farbigen Markierungen der Begriffe (gruen und weiss) helfen Dir bei der
    Zuordnung. Benutze wieder die Tasten 'W' und 'P' fuer die Sortierung zu den
    Bereichen nach links oder rechts. Berichtige wieder deine Fehler durch Druecken
    der anderen Taste."
94 </item>
95
96 <text spacebar>
97 / items = ("Zum Starten die Leertaste druecken!")
98 / position = (50%, 95%)
99 / valign = bottom
100 </text>
101
102 <text attributeA>
103 / items = attributeA
104 / fontstyle = ("Arial", 5%)
105 / txcolor = (0, 255, 0)
106 </text>
107
108 <text attributeB>
109 / items = attributeB
110 / fontstyle = ("Arial", 5%)
111 / txcolor = (0, 255, 0)
112 </text>
113
114 <text targetA>
115 / items = targetA
116 / fontstyle = ("Arial", 5%)
117 </text>
118
119 <text error>
120 / position = (50%, 75%)
121 / items = ("X")
122 / color = (255, 0, 0)
123 / fontstyle = ("Arial", 10%, true)
124 </text>
125
126 <text attributeAleft>
127 / items = attributeAlabel
128 / valign = top
129 / halign = left
130 / position = (5%, 5%)
131 / txcolor = (0, 255, 0)
132 / fontstyle = ("Arial", 5%)
133 </text>
134
135 <text attributeBright>
136 / items = attributeBlabel
```

```
137 / valign = top
138 / halign = right
139 / position = (95%, 5%)
140 / txcolor = (0, 255, 0)
141 / fontstyle = ("Arial", 5%)
142 </text>
143
144 <text targetAleftmixed>
145 / items = targetAlabel
146 / valign = top
147 / halign = left
148 / position = (5%, 19%)
149 / fontstyle = ("Arial", 5%)
150 </text>
151
152 <text targetArighmixed>
153 / items = targetAlabel
154 / valign = top
155 / halign = right
156 / position = (95%, 19%)
157 / fontstyle = ("Arial", 5%)
158 </text>
159
160 <text targetAleft>
161 / items = targetAlabel
162 / valign = top
163 / halign = left
164 / position = (5%, 5%)
165 / fontstyle = ("Arial", 5%)
166 </text>
167
168 <text targetArigh>
169 / items = targetAlabel
170 / valign = top
171 / halign = right
172 / position = (95%, 5%)
173 / fontstyle = ("Arial", 5%)
174 </text>
175
176 <text orleft>
177 / items = ("oder")
178 / valign = top
179 / halign = left
180 / position = (5%, 12%)
181 / fontstyle = ("Arial", 5%)
182 </text>
183
184 <text orright>
185 / items = ("oder")
186 / valign = top
187 / halign = right
188 / position = (95%, 12%)
189 / fontstyle = ("Arial", 5%)
190 </text>
191
192
193 <trial attributeA>
194 / validresponse = ("W", "P")
195 / correctresponse = ("W")
196 / stimulusframes = [1 = attributeA]
197 / posttrialpause = 250
198 </trial>
199
```

```

200 <trial attributeB>
201 / validresponse = ("W", "P")
202 / correctresponse = ("P")
203 / stimulusframes = [1 = attributeB]
204 / posttrialpause = 250
205 </trial>
206
207 <trial targetAleft>
208 / validresponse = ("W", "P")
209 / correctresponse = ("W")
210 / stimulusframes = [1 = targetA]
211 / posttrialpause = 250
212 </trial>
213
214 <trial targetAright>
215 / validresponse = ("W", "P")
216 / correctresponse = ("P")
217 / stimulusframes = [1 = targetA]
218 / posttrialpause = 250
219 </trial>
220
221
222 <block attributepractice>
223 / bgstim = (attributeAleft, attributeBright)
224 / trials = [1=instructions; 2-21 = random(attributeA, attributeB)]
225 / errormessage = true(error,200)
226 / responsemode = correct
227 </block>
228
229 <block compatibletest1>
230 / bgstim = (attributeAleft, orleft, targetAleftmixed, attributeBright)
231 / trials = [1=instructions; 2-21 = noreplace(targetAleft, targetAleft, attributeA,
232 / errormessage = true(error,200)
233 / responsemode = correct
234 / ontrialend = [if(block.compatibletest1.latency <= 10000 && block.compatibletest1.
235 / currenttrialnumber != 1) values.sum1a = values.sum1a + block.compatibletest1.
236 / latency]
237 / ontrialend = [if(block.compatibletest1.latency <= 10000 && block.compatibletest1.
238 / currenttrialnumber != 1) values.n1a = values.n1a + 1]
239 / ontrialend = [if(block.compatibletest1.latency <= 10000 && block.compatibletest1.
240 / currenttrialnumber != 1) values.ss1a = values.ss1a + (block.compatibletest1.
241 / latency * block.compatibletest1.latency)]
242 </block>
243
244 <block compatibletest2>
245 / bgstim = (attributeAleft, orleft, targetAleftmixed, attributeBright)
246 / trials = [1-35 = noreplace(targetAleft, targetAleft, attributeA, attributeA,
247 / attributeB, attributeB, attributeB)]
248 / errormessage = true(error,200)
249 / responsemode = correct
250 / ontrialend = [if(block.compatibletest2.latency <= 10000) values.sum1b = values.
251 / sum1b + block.compatibletest2.latency]
252 / ontrialend = [if(block.compatibletest2.latency <= 10000) values.n1b = values.n1b
253 / + 1]
254 / ontrialend = [if(block.compatibletest2.latency <= 10000) values.ss1b = values.
255 / ss1b + (block.compatibletest2.latency * block.compatibletest2.latency)]
256 </block>
257
258 <block incompatibletest1>
259 / bgstim = (attributeAleft, targetARightMixed, orright, attributeBright)

```

```

252 / trials = [1=instructions; 2-21 = noreplace(targetAright , targetAright , attributeA ,
      attributeA , attributeA , attributeB , attributeB)]
253 / errormessage = true(error,200)
254 / responsemode = correct
255 / ontrialend = [if(block.incompatibletest1.latency <= 10000 && block.
      incompatibletest1.currenttrialnumber != 1) values.sum2a = values.sum2a + block.
      incompatibletest1.latency]
256 / ontrialend = [if(block.incompatibletest1.latency <= 10000 && block.
      incompatibletest1.currenttrialnumber != 1 ) values.n2a = values.n2a + 1]
257 / ontrialend = [if(block.incompatibletest1.latency <= 10000 && block.
      incompatibletest1.currenttrialnumber != 1 ) values.ss2a = values.ss2a + (block.
      incompatibletest1.latency * block.incompatibletest1.latency)]
258 </block>
259
260 <block incompatibletest2>
261 / bgstim = (attributeAleft , targetArightMixed , orright , attributeBright)
262 / trials = [1-35 = noreplace(targetAright , targetAright , attributeA , attributeA ,
      attributeA , attributeB , attributeB)]
263 / errormessage = true(error,200)
264 / responsemode = correct
265 / ontrialend = [if(block.incompatibletest2.latency <= 10000) values.sum2b = values.
      sum2b + block.incompatibletest2.latency]
266 / ontrialend = [if(block.incompatibletest2.latency <= 10000) values.n2b = values.
      n2b + 1]
267 / ontrialend = [if(block.incompatibletest2.latency <= 10000) values.ss2b = values.
      ss2b + (block.incompatibletest2.latency * block.incompatibletest2.latency)]
268 </block>
269
270 <block compatibletestinstructions>
271 / bgstim = (attributeAleft , orleft , targetAleftmixed , attributeBright)
272 / trials = [1=instructions]
273 / recorddata = false
274 </block>
275
276 <block incompatibletestinstructions>
277 / bgstim = (attributeAleft , attributeBright , orright , targetArightmixed)
278 / trials = [1=instructions]
279 / recorddata = false
280 </block>
281
282 <block summary>
283 / trials = [1=summary]
284 / recorddata = false
285 </block>
286
287
288 <defaults>
289 / fontstyle = ("Arial", 3.5%)
290 / screencolor = (0,0,0)
291 / txbgcolor = (0,0,0)
292 / txcolor = (255, 255, 255)
293 / minimumversion = "3.0.0.0"
294 </defaults>
295
296
297 <expt>
298 / blocks = [1=attributepractice; 2=block1; 3=block3; 4=block4; 5=block5; 6=block6;
      7=block7]
299 </expt>
300
301 <variables>

```

```

302 / group = (1 of 2) (block1=compatibletest1; block3=compatibletestinstructions;
      block4=compatibletest2; block5=incompatibletest1; block6=
      incompatibletestinstructions; block7=incompatibletest2)
303 / group = (2 of 2) (block1=incompatibletest1; block3=incompatibletestinstructions;
      block4=incompatibletest2; block5=compatibletest1; block6=
      compatibletestinstructions; block7=compatibletest2)
304 </variables>
305
306
307 <data>
308 / columns = [date, time, subject, blockcode, blocknum, trialcode, trialnum, response
      , correct, latency, stimulusnumber, stimulusitem, expressions.da, expressions.db
      , expressions.d ]
309 </data>
310
311
312 <monkey>
313 / latencydistribution = normal(500, 100)
314 / percentcorrect = 90
315 </monkey>
316
317
318 <values>
319 / sum1a = 0
320 / sum2a = 0
321 / sum1b = 0
322 / sum2b = 0
323 / n1a = 0
324 / n2a = 0
325 / n1b = 0
326 / n2b = 0
327 / ss1a = 0
328 / ss2a = 0
329 / ss1b = 0
330 / ss2b = 0
331 / magnitude = "unknown"
332 </values>
333
334 * 1 is compatible, 2 is incompatible
335 * a is first block, b is second block
336
337 <expressions>
338 / m1a = values.sum1a / values.n1a
339 / m2a = values.sum2a / values.n2a
340 / m1b = values.sum1b / values.n1b
341 / m2b = values.sum2b / values.n2b
342 / sd1a = sqrt(((values.ss1a - (values.n1a * (expressions.m1a * expressions.m1a))) / (
      values.n1a - 1))
343 / sd2a = sqrt(((values.ss2a - (values.n2a * (expressions.m2a * expressions.m2a))) / (
      values.n2a - 1))
344 / sd1b = sqrt(((values.ss1b - (values.n1b * (expressions.m1b * expressions.m1b))) / (
      values.n1b - 1))
345 / sd2b = sqrt(((values.ss2b - (values.n2b * (expressions.m2b * expressions.m2b))) / (
      values.n2b - 1))
346 / sda = sqrt((((values.n1a - 1) * (expressions.sd1a * expressions.sd1a) + (values.
      n2a - 1) * (expressions.sd2a * expressions.sd2a)) + ((values.n1a + values.n2a) *
      ((expressions.m1a - expressions.m2a) * (expressions.m1a - expressions.m2a)) /
      4) ) / (values.n1a + values.n2a - 1) )
347 / sdb = sqrt((((values.n1b - 1) * (expressions.sd1b * expressions.sd1b) + (values.
      n2b - 1) * (expressions.sd2b * expressions.sd2b)) + ((values.n1b + values.n2b) *
      ((expressions.m1b - expressions.m2b) * (expressions.m1b - expressions.m2b)) /
      4) ) / (values.n1b + values.n2b - 1) )
348 / da = (m2a - m1a) / expressions.sda

```

```

349 / db = (m2b - m1b) / expressions.sdb
350 / d = (expressions.da + expressions.db) / 2
351 / attitude = "unknown"
352 </expressions>

```

SPSS-Skript für den *D*-Wert-Algorithmus

```

1 GET TRANSLATE FILE='Schwierigkeit.dat'
2 /TYPE=TAB /MAP /FIELDNAMES .
3
4 VALUE LABELS blocknum
5 1 'Target practice'
6 2 'Attribute practice'
7 3 'First pairing practice'
8 5 'First pairing test'
9 6 'Reversed target practice'
10 7 'Second pairing practice'
11 9 'Second pairing test' .
12
13 IF (MOD(subject,2) = 1) ORDER = 1 .
14 IF (MOD(subject,2) = 0) ORDER = 2 .
15
16 COMPUTE PAIRING = 0.
17 IF ((ORDER=1) and (blocknum=3|blocknum=5)) PAIRING = 1.
18 IF ((ORDER=1) and (blocknum=7|blocknum=9)) PAIRING = 2.
19 IF ((ORDER=2) and (blocknum=7|blocknum=9)) PAIRING = 1.
20 IF ((ORDER=2) and (blocknum=3|blocknum=5)) PAIRING = 2.
21
22 COMPUTE TEST = 0.
23 IF (blocknum=3|blocknum=7) TEST = 1.
24 IF (blocknum=5|blocknum=9) TEST = 2.
25
26 VALUE LABELS
27 TEST 0 'single-task practice' 1 '1st combined block' 2 '2nd combined block'
28 / correct 0 'error' 1 'correct' .
29
30 VARIABLE LABELS
31 correct "0=error, 1=correct"
32 / blocknum "block number"
33 / trialnum "trial number"
34 / ORDER "order of combined tasks"
35 / PAIRING "paired categories"
36 / TEST "1st or 2nd combined block".
37
38 VALUE LABELS
39 PAIRING
40 0 'single task practice'
41 1 'compatible'
42 2 'incompatible'
43 / ORDER
44 1 'compatible first'
45 2 'incompatible first'.
46 COMPUTE error = 100*(1 - correct).
47
48 * Save data to re-use for computing other measures .
49 SAVE OUTFILE = 'Schwierigkeit_temp.sav' .
50
51 DESCRIPTIVES ALL .
52
53 SUMMARIZE TABLES latency BY TEST BY PAIRING.
54

```

```

55 GET FILE = 'Schwierigkeit_temp.sav' .
56 * Use data for Blocks 3,5,7, and 9 (i.e., TEST = 1 or 2) .
57 SELECT IF (TEST=1 OR TEST=2) .
58
59 * Mark trials with latency < 300 ms (for possible subject discard) .
60 COMPUTE FLAG_300 = 0 .
61 IF (LATENCY < 300) FLAG_300 = 1 .
62 * Mark trials with latency < 400 ms (to provide a count of these) .
63 COMPUTE FLAG_400 = 0 .
64 IF (LATENCY < 400) FLAG_400 = 1 .
65 * Mark trials with latency > 10000 ms (to provide a count of these) .
66 COMPUTE FLAG_10K = 0 .
67 IF (LATENCY GT 10000) FLAG_10K = 1 .
68
69 * Following line can be used to check frequencies of fast & slow responses & errors
70 FREQUENCIES FLAG_300 FLAG_400 FLAG_10K ERROR .
71
72 * Record criteria for potential use in subject discards .
73 * This count can later be used as the basis for subject discard .
74 AGGREGATE OUTFILE = 'CRITERIA.SAV'
75 / BREAK = SUBJECT ORDER
76 / PCT_300 = PGT(FLAG_300,0)
77 / PCT_400 = PGT(FLAG_400,0)
78 / PCT_10K = PGT(FLAG_10K,0)
79 / AVELTNCY ERRORPCT = MEAN (LATENCY ERROR)
80 / NTRIALS = N .
81
82 * Drop trials slower than 10000 ms for LATENCY .
83 SELECT IF (LATENCY LE 10000) .
84
85 * The following line deletes latencies less than 300. To prevent these trials from
    being
86 * discarded, add an asterisk before the next line to comment it out.
87 * SELECT IF (LATENCY GE 300) .
88
89 DESCRIPTIVES ALL.
90
91 * The following aggregate command computes block means and SDs needed for the D
    measures.
92 AGGREGATE OUTFILE = *
93 / BREAK = SUBJECT BLOCKNUM ORDER PAIRING TEST
94 / MEAN_LAT MEAN_ERR = MEAN(latency error) / SD_LAT = SD(latency)
95 / NTRIALS = N .
96
97 DESCRIPTIVES ALL.
98
99 *Save this as a record of this stage of analysis .
100 SAVE OUTFILE = 'Schwierigkeit_BIEP.ALL_MEANS&SDs.SAV' .
101
102 IF (PAIRING = 1) N1 = NTRIALS .
103 IF (PAIRING = 2) N2 = NTRIALS .
104 IF (PAIRING = 1) M1 = MEAN_LAT .
105 IF (PAIRING = 2) M2 = MEAN_LAT .
106 IF (PAIRING = 1) ERR1 = MEAN_ERR .
107 IF (PAIRING = 2) ERR2 = MEAN_ERR .
108 IF (PAIRING = 1) SD1 = SD_LAT .
109 IF (PAIRING = 2) SD2 = SD_LAT .
110
111 AGGREGATE OUTFILE = *
112 / BREAK = SUBJECT ORDER TEST
113 / M1 M2 ERR1 ERR2 N1 N2 SD1 SD2 =
114 FIRST(M1 M2 ERR1 ERR2 N1 N2 SD1 SD2) .

```



```

115
116 *Save this as a record of this stage of analysis .
117 SAVE OUTFILE = 'Schwierigkeit_BIEP.MEANS&SDS.SAV' .
118
119 IF (TEST=1) ERR1a = ERR1 .
120 IF (TEST=1) ERR2a = ERR2 .
121 IF (TEST=2) ERR1b = ERR1 .
122 IF (TEST=2) ERR2b = ERR2 .
123
124 * These are the numerator components in millisecond units .
125 IF (TEST=1) M1a = M1 .
126 IF (TEST=1) M2a = M2 .
127 IF (TEST=2) M1b = M1 .
128 IF (TEST=2) M2b = M2 .
129
130 IF (TEST=1) SD1a = SD1 .
131 IF (TEST=1) SD2a = SD2 .
132 IF (TEST=2) SD1b = SD1 .
133 IF (TEST=2) SD2b = SD2 .
134
135 IF (TEST=1) N1a = N1 .
136 IF (TEST=1) N2a = N2 .
137 IF (TEST=2) N1b = N1 .
138 IF (TEST=2) N2b = N2 .
139
140 COMPUTE D_asis_num = M2 - M1.
141
142 *Use SD based on all responses (StanDevX) as denominator for D_biep .
143 COMPUTE D_asis_denom = SQRT( ( ((N1-1) * SD1**2 + (N2-1) * SD2**2)
144 + ((N1+N2) * ((M1-M2)**2) / 4) ) / (N1 + N2 - 1) ) .
145
146 IF (TEST=1) D_BIEP_a = D_asis_num / D_asis_denom .
147 IF (TEST=2) D_BIEP_b = D_asis_num / D_asis_denom .
148 IF (TEST=1) Na = N1+N2 .
149 IF (TEST=2) Nb = N1+N2 .
150
151 DESCRIPTIVES ALL .
152
153 AGGREGATE OUTFILE = *
154 / BREAK = SUBJECT ORDER
155 / D_biep_a D_biep_b M1a M2a M1b M2b ERR1a ERR2a ERR1b ERR2b Na Nb =
156 FIRST(D_biep_a D_biep_b M1a M2a M1b M2b ERR1a ERR2a ERR1b ERR2b Na Nb) .
157
158 VARIABLE LABELS
159 D_biep_a 'd score 1st blocks'
160 D_biep_b 'd score 2nd blocks'
161 Na '# trials, 1st combined blocks'
162 Nb '# trials, 2nd combined blocks'
163 M1a 'Mn Lat. 1st block, pairing 1'
164 M2a 'Mn Lat. 1st block, pairing 2'
165 M1b 'Mn Lat. 2nd block, pairing 1'
166 M2b 'Mn Lat. 2nd block, pairing 2'
167 ERR1a 'Error % 1st block, pairing 1'
168 ERR2a 'Error % 1st block, pairing 2'
169 ERR1b 'Error % 2nd block, pairing 1'
170 ERR2b 'Error % 2nd block, pairing 2'
171 / SUBJECT 'Session ID' .
172
173 COMPUTE ERR_1 = (ERR1a + ERR1b) / 2 .
174 COMPUTE ERR_2 = (ERR2a + ERR2b) / 2 .
175
176 VARIABLE LABELS
177 ERR_1 'Error % for Pairing 1 (both combined tasks)'

```

```

178 / ERR_2 'Error % for Pairing 2 (both combined tasks)' .
179
180 *Compute unweighted combination of measures based on first and second blocks
181 of combined tasks.
182 * !! Do this even if there are fewer trials in the first block because the
183 * first block has been found to have as good or better validity, even with
184 * fewer trials (see Greenwald, Nosek, & Banaji, JPSP 2003) .
185
186 COMPUTE D_biep = (D_biep_a + D_biep_b) / 2 .
187 VARIABLE LABELS
188     D_biep 'd score all blocks' .
189
190
191 *****
192 ** TO REVERSE THE SCORING OF THE IAT, REMOVE THE ASTERISKS **
193 ** FROM THE NEXT FOUR COMMAND LINES. **
194 ***** .
195
196 * COMPUTE D_biep = 0 - D_biep .
197 * COMPUTE D_biep_a = 0 - D_biep_a .
198 * COMPUTE D_biep_b = 0 - D_biep_b .
199 * VARIABLE LABELS
200     D_biep 'd score all blocks'
201     / D_biep_a 'd score 1st blocks'
202     / D_biep_b 'd score 2nd blocks' .
203
204 FORMAT D_biep D_BIEP_a D_BIEP_b D_BIEP (F6.3) SUBJECT (F7.0)
205     Na Nb (F4.0) M1a M2a M1b M2b (F7.1)
206     ERR1a ERR2a ERR1b ERR2b ERR_1 ERR_2 (F5.1) .
207
208 SAVE OUTFILE = 'Schwierigkeit_D_BIEP_temp.SAV' .
209 DESCRIPTIVES ALL .
210
211 *****
212 **** STEP 3: COMBINE MEASURES AND SAVE ****
213 *****
214
215 *Add labels to CRITERIA.SAV .
216 GET FILE = 'Schwierigkeit_CRITERIA.SAV' .
217 VARIABLE LABELS
218     PCT_300 '% latencies below 300 ms'
219     / PCT_400 '% latencies below 400 ms'
220     / PCT_10K '% latencies above 10,000 ms'
221     / AVELTNCY 'Overall average latency (ms)'
222     / ERRORPCT 'Overall % errors'
223     / NTRIALS 'Total # of combined-task trials' .
224
225 FORMAT AVELTNCY (F7.1) PCT_300 PCT_400 PCT_10K ERRORPCT (F6.1) SUBJECT (F7.0) .
226
227 SAVE OUTFILE = 'Schwierigkeit_CRITERIA.SAV' .
228
229 *Combine and save .
230 MATCH FILES FILE = 'Schwierigkeit_D_BIEP_temp.SAV' / FILE = 'Schwierigkeit_CRITERIA.
    SAV' /BY SUBJECT .
231
232 SAVE OUTFILE = 'Schwierigkeit_D_BIEP.SAV' .
233
234 * Note: No subjects have yet been discarded. However, each record contains .
235 * three variables that might be used for discard .
236 * PCT_300 = overall % of responses with latencies below 300 ms .
237 * ERRORPCT = overall error % .
238 * AVELTNCY = overall average latency .
239 * [Note that 'aveltny' was computed BEFORE discard of latencies above 10000 ms] .

```

```
240
241 GET FILE = 'Schwierigkeit_D_BIEP.SAV' .
242 DESCRIPTIVES ALL .
```

A.2 Interessefragebogen

Interessierst du dich für Physik?

Eine Befragung der Humboldt-Universität zu Berlin
zum Fachinteresse in Physik



Liebe Schülerin, lieber Schüler,

Mithilfe dieser Befragung möchte ich herausfinden, wie sehr du dich für Physik interessierst. Hier gibt es keine richtigen oder falschen Antworten, du kannst also keine Fehler machen! Deine Antworten sind mir jedoch sehr wichtig. Daher bitte ich dich, alle Fragen gewissenhaft zu beantworten.

Vielen Dank für Deine Unterstützung!

Nico Westphal

Befragung zum Fachinteresse in Physik



Hinweise zum Ausfüllen

Bei den folgenden Fragen sollst du **das für dich richtige Kästchen** ankreuzen. Bitte kreuze bei jeder Frage, also in jeder Zeile, nur **ein** Kästchen an!

Wenn du mal versehentlich nicht das Kästchen angekreuzt hast, welches du eigentlich wählen wolltest, dann kannst du deine Antwort korrigieren.

Beispiel:

Wie sehr stimmst du folgenden Aussagen zu?
(Bitte in jeder Zeile nur ein Kästchen ankreuzen.)

	In fast keiner Unterrichtsstunde	In einigen Unterrichtsstunden	In den meisten Unterrichtsstunden	In jeder Unterrichtsstunde	kann ich nicht beantworten
Der Lehrer schreibt so undeutlich, dass ich an der Tafel nichts erkennen kann.	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

ungültig gemachte Antwort deine Antwort

Wie sehr stimmst du folgenden Aussagen zu?
(Bitte in jeder Zeile nur ein Kästchen ankreuzen.)

	In fast keiner Unterrichtsstunde	In einigen Unterrichtsstunden	In den meisten Unterrichtsstunden	In jeder Unterrichtsstunde	kann ich nicht beantworten
1 Im Unterricht habe ich den Eindruck, dass mich die Lehrerin / der Lehrer wahrnimmt.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2 Die Lehrerin/ der Lehrer schätzt meine Mitarbeit im Unterricht.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3 Im Unterricht habe ich das Gefühl, dass ich dem Lehrer / der Lehrerin wichtig bin.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4 Die Lehrerin/ der Lehrer traut mir zu, dass ich die naturwissenschaftlichen Inhalte verstehen kann.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5 Die Lehrerin/ der Lehrer traut mir anspruchsvolle Dinge zu.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6 Der Lehrer/ die Lehrerin traut uns zu, dass wir selbständig arbeiten.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7 Die Lehrerin/ der Lehrer ist offen für unterschiedliche Beiträge der Schüler.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8 Die Lehrerin/ der Lehrer ermuntert uns, selbst zu überlegen, wie man am besten vorgeht.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9 Der Lehrer/ die Lehrerin gibt uns genügend Möglichkeiten, selbständig zu arbeiten.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Befragung zum Fachinteresse in Physik


Wie sehr stimmst du mit den folgenden Aussagen überein?
(Bitte in jeder Zeile nur ein Kästchen ankreuzen.)

		stimme ganz zu	stimme eher zu	stimme eher nicht zu	stimme gar nicht zu	kann ich nicht beantworten
10	Im Allgemeinen macht es mir Spaß, mich mit naturwissenschaftlichen Themen zu befassen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11	Ich lese gerne etwas über Naturwissenschaften.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
12	Ich beschäftige mich gerne mit naturwissenschaftlichen Problemen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
13	Ich eigne mir gerne neues Wissen in den Naturwissenschaften an.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
14	Ich bin interessiert, Neues in den Naturwissenschaften zu lernen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
15	Sich in Physik anzustrengen, zahlt sich aus, weil mir das bei der Arbeit, die ich später machen möchte, helfen wird.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
16	Was ich in Physik lerne, ist wichtig für mich, weil ich es für meine spätere Ausbildung oder für mein Studium brauche.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
17	Ich lerne Physik, weil ich weiß, dass es für mich nützlich ist.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
18	Physik zu lernen, lohnt sich für mich, weil das Gelernte meine beruflichen Aussichten verbessern wird.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
19	Ich werde in Physik viele Dinge lernen, die mir helfen werden, einen Job zu bekommen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Wie hoch ist dein Interesse, etwas über die folgenden naturwissenschaftlichen Themenbereiche zu lernen?
(Bitte in jeder Zeile nur ein Kästchen ankreuzen.)

		hohes Interesse	durchschnittliches Interesse	geringes Interesse	kein Interesse	kann ich nicht beantworten
20	Themenbereiche in der Physik.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
21	Themenbereiche in der Chemie.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
22	Die Biologie der Pflanzen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
23	Die Biologie des Menschen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
24	Themenbereiche in der Astronomie.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
25	Themenbereiche in den Erdwissenschaften (Erdkunde, Geografie).	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
26	Wie Naturwissenschaftlerinnen und -wissenschaftler Experimente gestalten.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
27	Was für naturwissenschaftliche Erklärungen erforderlich ist.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Befragung zum Fachinteresse in Physik


Im Unterricht in Physik...
(Bitte in jeder Zeile nur ein Kästchen ankreuzen.)

	fast immer	oft	manchmal	nie	kann ich nicht beantworten
28 habe ich keine Lust mich zu beteiligen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
29 habe ich keine Lust, mich mit den Inhalten auseinander zu setzen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
30 bin ich mit meinen Gedanken woanders.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
31 tue ich nur so viel, wie von mir verlangt wird.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
32 arbeite ich nur mit, wenn ich dazu aufgefordert werde.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
33 beteilige ich mich nur, wenn es nicht anders geht.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
34 arbeite ich mit, weil ich es immer so mache.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
35 strengere ich mich an, weil ich gerne alles richtig mache.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
36 beteilige ich mich, weil es sich für eine Schülerin / einen Schüler gehört.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
37 arbeite ich mit, weil ich die Inhalte später bestimmt gebrauchen kann.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
38 sind mir die Inhalte wichtig, weil sie auch für viele Dinge im Alltag sinnvoll sind.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
39 arbeite ich mit, damit ich mich später in diesem Fach auskenne.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
40 finde ich die Inhalte richtig spannend.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
41 bin ich mit Freude dabei.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
42 macht mir der Unterricht Spaß.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
43 möchte ich gerne mehr über dieses Thema erfahren.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
44 bekomme ich Lust, mich weiter damit zu beschäftigen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
45 wird deutlich, dass der Lernstoff auch im Alltag wichtig ist.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
46 wird erkennbar, wofür die Experimente nützlich sind.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
47 lernt man, wo die Inhalte praktisch angewendet werden.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
48 werden wir darüber informiert, dass dieser Stoff wichtig für andere Themen ist.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Vielen Dank für deine Mitarbeit!

ANHANG B

Unterrichtsmaterialien

Folgend werden die im Treatment verwendeten Unterrichtsmaterialien aufgeführt. Sie beinhalten Arbeitsblätter, Übungsaufgaben und Aufgabenblätter an Experimentierstationen. Die dargestellten Materialien sind teilweise skaliert und an den Rändern beschnitten.

B.1 Arbeitsblatt zum Prinzip Ameise

Arbeitsblatt:
Das Prinzip Ameise

Aufgaben:
Teil A: Schreibe in die Sprechblasen, was die Ameise sieht.

Weg der Ameise

Punktlicht-
leuchte

Schattengeber

Teil B: Was sieht die Ameise? Überlegt gemeinsam und zeichnet Hilfslinien ein. Schreibt die Antworten dann in die Sprechblasen.

Weg der Ameise

Stablampe

Schattengeber

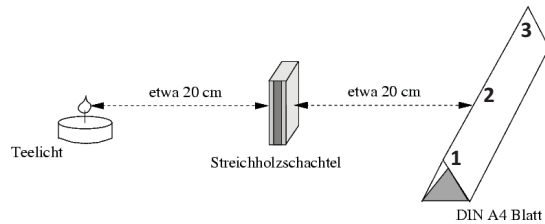
B.2 Aufgabenblatt für das Schülerexperiment Schattenbereiche

Arbeitsblatt:

Schülerversuch. Thema: Schatten

Material: 2 Teelichter, 1 Streichholzschachtel, 1 weißes DIN A4 Blatt

Aufbau 1: *Vorsicht beim Umgang mit den Teelichtern! Der Tisch muss bis auf die Versuchsgüter frei bleiben!*

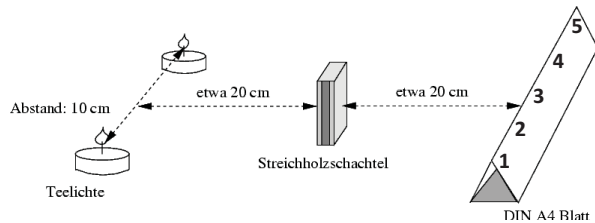


Durchführung: Der Raum muss etwas abgedunkelt werden. Zünde die Kerzen an und beobachte die Schatten auf dem Papier!

Aufgaben: Notiere die Antworten zu den Fragen!

1. Was sieht die Ameise, die auf dem Papier entlangkrabbelt, wenn sie vom Bereich des Schattens (2) und von den Bereichen rechts und links neben dem Schatten (1 & 3) zu der Kerze schaut?
2. Vergleiche die Helligkeit auf dem Papier mit dem, was die Ameise sieht!

Aufbau 2:



Aufgaben: Notiere die Antworten zu den Fragen!

1. Was sieht die Ameise, die auf dem Papier entlangkrabbelt, wenn sie von den Bereichen der Schatten (2 & 4) und von den Bereichen daneben und dazwischen (1, 3 & 5) zu der Kerze schaut?
2. Rücke das Papier langsam bis auf 5 cm an die Schachtel heran!
 - a) Notiere die Veränderung der Schatten!
 - b) Was sieht nun die Ameise, wenn sie von den verschiedenen Bereichen zu der Kerze schaut?
 - c) Vergleiche die Helligkeit auf dem Papier mit dem, was die Ameise sieht!

B.3 Arbeitsblatt zu den Schattenformen ausgedehnter Lichtquellen

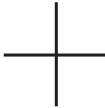




Arbeitsblatt:

Verschiedene Schattenformen

Aufgabe:

Zeichne den Aufbau von der Seite.

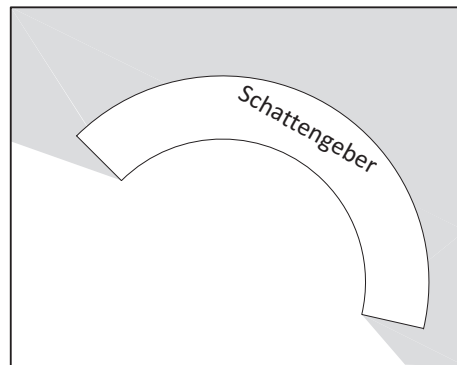
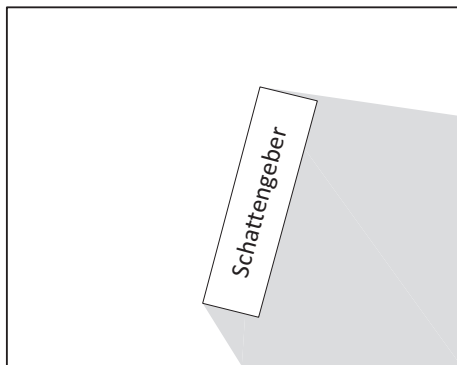
Beobachte die Experimente mit den verschiedenen Lichtquellen und den verschiedenen Schattengebern genau. Zeichne in die Felder der Tabelle ein, wie der Schatten jeweils aussieht, wenn man einen der oben abgebildeten Schattenkörper mit einer der links abgebildeten Lichtquelle beleuchtet.

<div>Schattengeber</div> <div>Lichtquelle</div>			
<div>⊗</div> <div>Punktlichte</div>	Ich sehe folgenden Schatten:	Ich sehe folgenden Schatten:	Ich sehe folgenden Schatten:
 <div>Stableuchte quer</div>	Ich sehe folgenden Schatten:	Ich sehe folgenden Schatten:	Ich sehe folgenden Schatten:
 <div>Stableuchte hochkant</div>	Ich sehe folgenden Schatten:	Ich sehe folgenden Schatten:	Ich sehe folgenden Schatten:
<div>Wie sieht diese Lichtquelle aus?</div>	Ich sehe folgenden Schatten:	Ich sehe folgenden Schatten:	Ich sehe folgenden Schatten:

B.4 Übungen zu Schatten, Sicht- und Lichtwegen

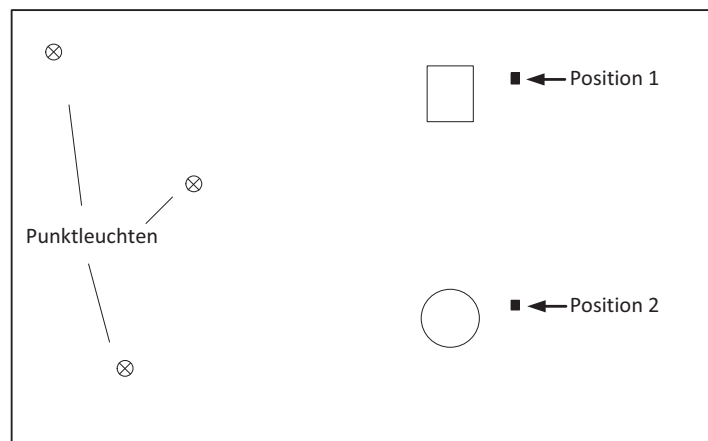
1. Aufgabe

Du schaust von oben auf die abgebildeten Schattengeber und siehst auf dem Fußboden die eingezeichneten Schatten. Zeichne den Ort ein, an dem die Leuchte stehen muss, welche die Schatten erzeugt!



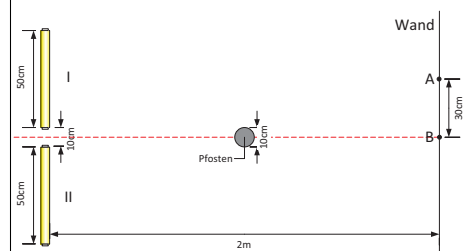
2. Aufgabe

Zeichne alle Schattenbereiche ein!



Ist es an Position 1 oder an Position 2 heller? Begründe mit unseren Merksätzen deine Antwort!

3. Aufgabe



Zwei 50cm lange Leuchtstofflampen I und II stehen wie im Bild vor einer Wand. Zwischen der Wand und den Lampen befindet sich ein 10 cm dicker Pfosten.

- Fertige eine Zeichnung im Maßstab 1:10 an.
- Bestimme, wie viel Zentimeter Leuchtstoffröhre von den Punkten A und B aus jeweils zu sehen sind!
- An welchem Ort ist es heller?

B.5 Aufgabenblatt für Schülerexperimente zum 1. Spiegelgesetz

Arbeitsblatt:

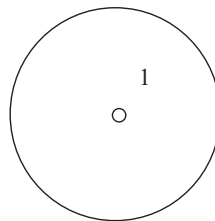
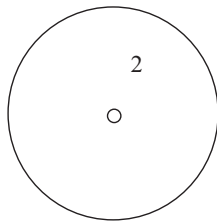
Schülerversuch: Wo muss der Spiegel stehen?

Material: 2 Teelichter, 1 Spiegelkachel, 1 Ständer für die Spiegelkachel

Aufgaben:

Stell den Spiegel so auf, dass das Spiegelbild von Kerze 1 mit der „wirklich vorhandenen“ Kerze 2 zusammenfällt.

Zeichne die Lage des Spiegels als Linie ein! Zeichne auch die Verbindungslinie zwischen den Kerzen. Wie liegen die beiden Linien zueinander?



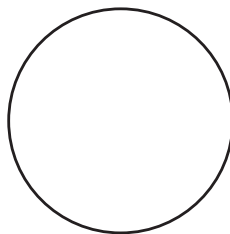
Arbeitsblatt:

Schülerversuch: Wo befindet sich die Spiegelkerze?

Material: 1 Teelicht, 1 Spiegelkachel, 1 Ständer für die Spiegelkachel

Aufgabe:

Stell die Spiegelkachel auf der vorgezeichneten Linie und das Teelicht auf dem Kreis auf. Zünde das Teelicht nun an. Zeichne diejenigen Schatten nach, die von der Spiegelkerze am Rand des Spiegels erzeugt werden. Verlängere nun die Schattengrenzen der gezeichneten Schatten in Richtung der Spiegelkerze (Lichtwege!). Wo beginnen die Lichtwege? Wo ist die Spiegelkerze? Miss die Abstände zwischen Spiegel, Kerze und Spiegelkerze!



B.6 Arbeitsblatt zur Spiegelwelt

Arbeitsblatt:
Schatten in der Spiegelwelt

Skizze:

Aufgaben:

1. Trage deine Beobachtungen in die Skizze ein.
2. Ordne allen Schatten den richtigen Platz in der Tabelle zu, indem du die jeweilige Nummer aus dem Tafelbild einträgst.

<div style="display: inline-block; transform: rotate(-45deg); transform-origin: left top;">Lichtquelle</div> <div style="display: inline-block;">Schattenkörper</div>	Kerze	Spiegelkerze
Zylinder		
Spiegelzylinder		
Kerze		
Spiegelkerze		

B.7 Arbeitsblatt zum Doppelschattenexperiment

Das Doppelschattenexperiment

Aufgabe:

Trage die gespiegelte Hand und die gespiegelte Lichtquelle ein. Erkläre dann durch Einzeichnen, wie die beiden Schatten an der Decke entstehen.

Zimmerdecke



Hand


Spiegel

B.8 Lernerfolgskontrolle

Lernerfolgskontrolle

Thema: Optik – Licht, Sehen, Schatten, Spiegel



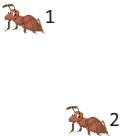



Name:

Datum:

Klasse:

Punkte: / 20

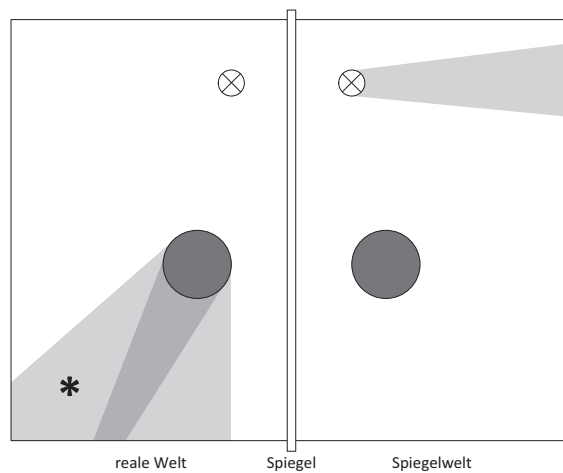
Zensur:

Aufgaben	Punkte
<p>1. Können die Ameisen die Stablampe sehen? Zeichne die Blickwege beider Ameisen ein und begründe deine Antwort in Satzform!</p> <div style="display: flex; align-items: center; justify-content: space-around;"> <div style="text-align: center;">  <p>Stablampe</p> </div> <div style="text-align: center;">  <p>Schattengeber</p> </div> <div style="text-align: center;">  </div> </div>	/ 3
<p>2. Die Fußballspieler erzeugen auf dem Rasen vier Schatten (siehe Abbildung).</p> <p>a) Wie viele Lichtquellen sind an der Erzeugung dieser Schattenbilder beteiligt?</p> <p>b) Erkläre das Zustandekommen dieser Schattenbilder!</p> <div style="text-align: center;">  </div>	/ 3
<p>3. Ein Schattengeber wird mit einer Stablampe beleuchtet. Welche Schattenform ist zu erwarten?</p> <div style="display: flex; align-items: center; justify-content: space-around;"> <div style="text-align: center;"> <p>Stablampe</p>  </div> <div style="text-align: center;"> <p>Schattengeber</p>  </div> </div> <p>Schattenform:</p>	/ 4

4. Donald ist leicht kurzsichtig. Dinge, die bis zu 1 m entfernt sind, kann er jedoch scharf erkennen. Sieht Donald sein Spiegelbild scharf, wenn er 1 m vor dem Spiegel steht? Begründe deine Antwort!

/ 2

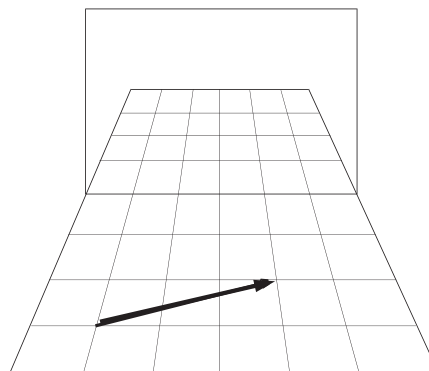
5. Vor einem Spiegel steht ein Zylinder aus Pappe, der von einer Halogenlampe beleuchtet wird. Die Skizze zeigt die Anordnung von oben:



- a) In der Skizze fehlen drei Schatten. Ergänze sie.
b) Erkläre, wie der mit dem Stern gekennzeichnete Schatten zustande kommt!

/ 5

6. Vor einem Spiegel hat jemand einen Pfeil auf den Boden gezeichnet. Trage in die Skizze ein, wie der Pfeil in der Spiegelwelt aussehen muss! Begründe mit Hilfe der Spiegelgesetze deine Entscheidung!



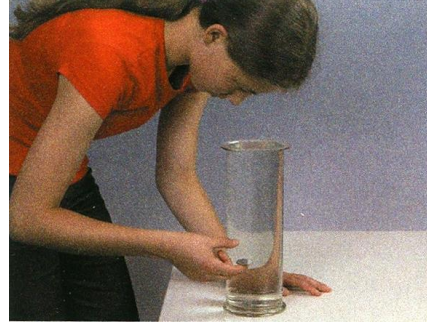
/ 3

Viel Erfolg!!!

B.9 Aufgabenblatt für die Experimentierstationen zur Hebung

Station: Senkrechte Hebung

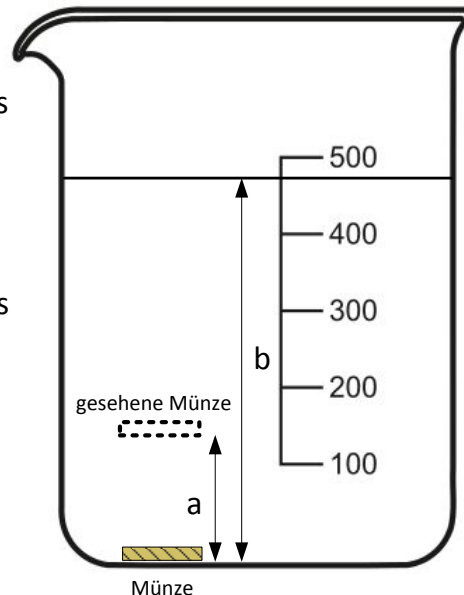
Schaue von oben in einen hohen wassergefüllten Becher, auf dessen Grund eine Münze liegt. Schätze den Ort der gesehenen Münze, indem du eine zweite Münze neben dem Becher in die entsprechende Höhe hältst!



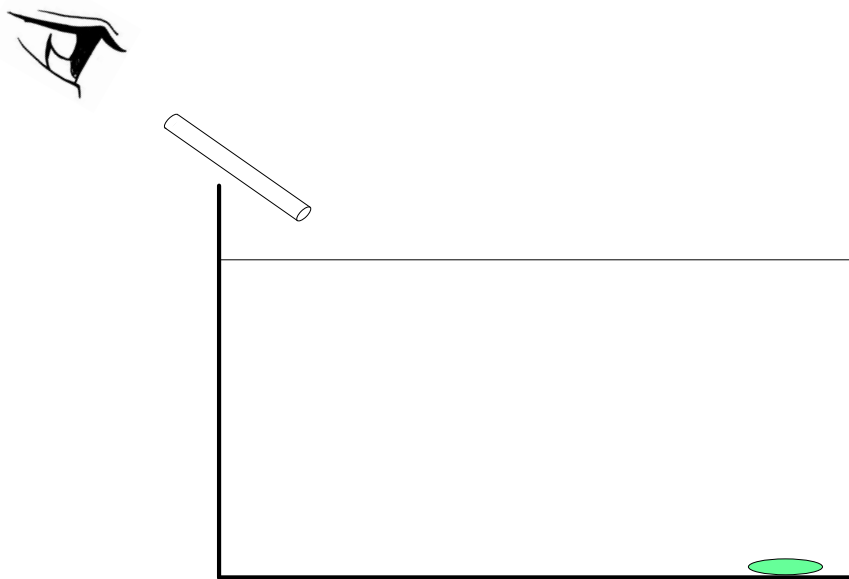
Miss die Höhe (a) vom Boden bis zu dem Ort, wo du die Münze siehst!

Vergleiche sie mit der Gesamthöhe des Wasserspiegels (b), indem du den Quotienten (a/b) berechnest!

Vergleiche deine Ergebnisse mit deinen Mitschülern!



Station: Harpunieren



Schaue von schräg oben in ein wassergefülltes Becken, auf dessen Grund ein Stein liegt!

Ziele mit dem Fernrohr auf den Stein!

Prüfe nun mit der „Harpune“ (Metallstange), ob du den Stein triffst, wenn du sie durch das Fernrohr steckst!

Skizziere den Aufbau und notiere deine Beobachtungen!

Grafiken, Tabellen und Formeln

Tabelle C.1 Verwendete Effektstärkemaße mit deren Einteilung in schwache, mittlere und starke Effekte (Cohen, 1992; Greenwald et al., 2003)

	Effektstärke		
	schwach	mittel	stark
<i>D</i> -Wert	0.15	0.35	0.50
Pearsons <i>r</i>	.10	.30	.50
Spearman's ρ	.10	.30	.50
Cohens <i>d</i>	0.20	0.50	0.80
η_p^2	.01	.06	.14

C.1 Testgüte und Skalierung

Kuder Richardson-Formel 8:

$$KR_8 = \frac{s_X^2 - \sum_i p_i q_i}{2 \cdot s_X^2} + \sqrt{\frac{\sum_i r_{it}^2 + p_i q_i}{s_X^2} + \left(\frac{s_X^2 - \sum_i p_i q_i}{2 \cdot s_X^2} \right)^2} \quad (C.1)$$

s_X^2 = Varianz der Test-Rohwerte

p = Schwierigkeitsindex

q = $(1 - p)$

r_{it} = Trennschärfekoeffizient

Tabelle C.2 Cronbachs α -Koeffizienten für alle IATs

	α	
	Vortest	Nachtest
Schwierigkeit	.79	.89
Geschlechtskonnotation	.93	.72
Selbstkonzept	.90	.82

Tabelle C.3 Cronbachs α -Koeffizienten für alle Interessesskalen

	α	
	Vortest	Nachtest
wse	.75	.76
wku	.73	.71
wau	.68	.60
wsewkuwau	.83	.83
JOYSCIE	.90	.89
INSTSCIE	.90	.90
INTSCIE	.72	.73
mo1	.77	.80
mo2	.82	.81
mo3	.92	.90
mo123	.92	.91
meco	.83	.78

Tabelle C.4 Itemkennwerte der Raschanalyse

Item	Parameter		Fit-Statistiken			
	Schwierigkeit	Fehler	Infit <i>MNSQ</i>	Outfit <i>MNSQ</i>	Infit <i>t</i>	Outfit <i>t</i>
1	1.41	0.17	1.06	1.11	0.81	1.04
2	-0.49	0.16	1.02	0.95	0.32	-0.33
3	-1.86	0.23	0.92	0.73	-0.43	-0.86
4	1.80	0.15	1.02	0.99	0.26	-0.04
5	-2.54	0.28	0.95	0.62	-0.15	-0.92
6	0.24	0.14	1.06	1.10	1.03	1.15
7	0.88	0.14	0.91	0.86	-1.85	-1.96
8	-0.57	0.16	1.15	1.34	1.82	2.29
9	0.22	0.15	0.98	0.92	-0.29	-0.82
10	2.12	0.17	0.95	0.85	-0.54	-1.00
11	-1.21	0.18	0.98	1.01	-0.16	0.13

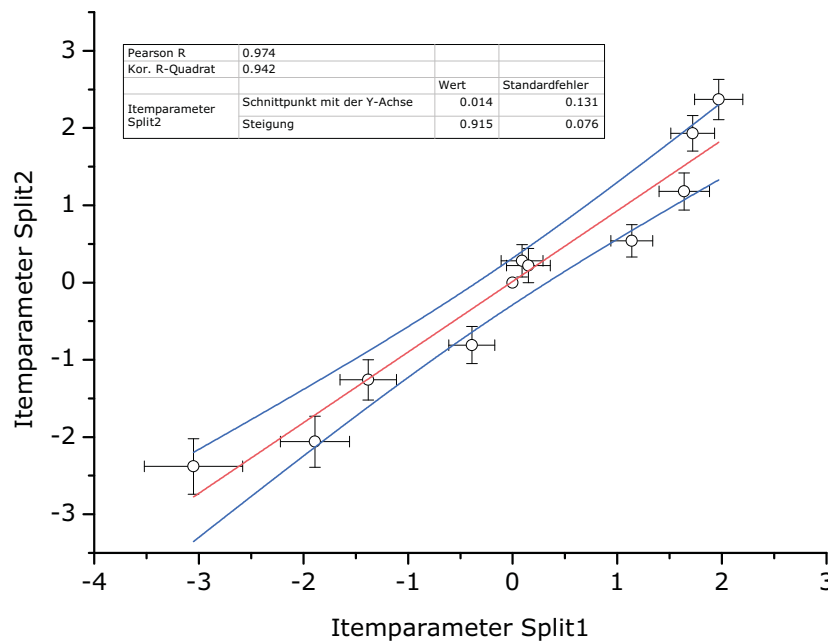


Abb. C.1 Split-Half-Analyse der Itemparameter mit Regressionsgerade und 95 %-Konfidenzintervall

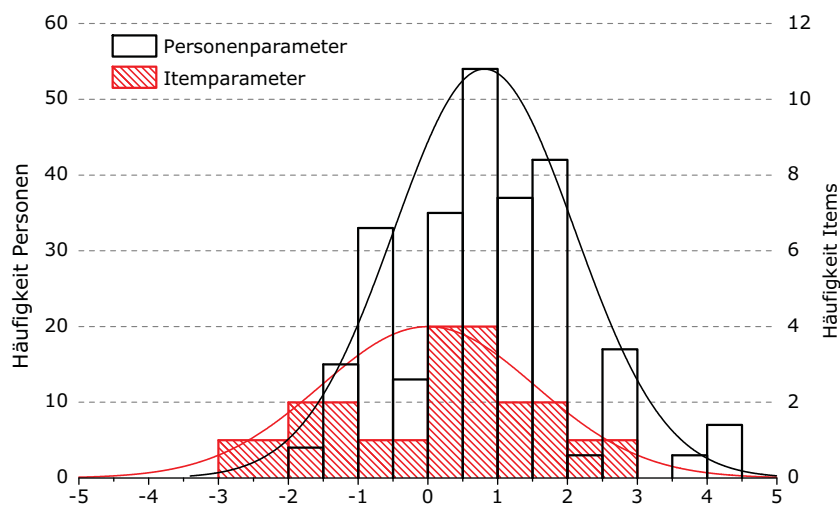


Abb. C.2 Überdeckung der Personen- und Itemparameterverteilungen. Die Verteilung der Personenparameter ist leicht nach rechts verschoben ($M = 0.8$), was bedeutet, dass die Items für die Stichprobe etwas zu leicht waren.

C.2 Tests auf Normalverteilung

Tabelle C.5 Normalverteilungstests aller *D*-Werte im Vor- und Nachtest

			Kolmogorov-Smirnov			Shapiro-Wilk		
			Stat.	<i>df</i>	Sig.	Stat.	<i>df</i>	Sig.
Vortest	Schwierigkeit	KG	.058	128	.200	.995	128	.931
		VG	.051	105	.200	.990	105	.643
	Geschlecht	KG	.114	128	.000	.964	128	.002
		VG	.055	105	.200	.993	105	.884
	Selbstkonzept	KG	.055	128	.200	.990	128	.447
		VG	.047	105	.200	.982	105	.177
Nachtest	Schwierigkeit	KG	.066	111	.200	.986	111	.307
		VG	.064	100	.200	.980	100	.124
	Geschlecht	KG	.041	111	.200	.994	111	.901
		VG	.042	100	.200	.989	100	.571
	Selbstkonzept	KG	.078	111	.097	.987	111	.341
		VG	.045	100	.200	.991	100	.736

Anmerkung Signifikanzen ($p < .05$) sind fett hervorgehoben.

Tabelle C.6 Normalverteilungstests aller Mittelwerte der semantischen Differentiale im Vor- und Nachtest

			Kolmogorov-Smirnov			Shapiro-Wilk		
			Stat.	<i>df</i>	Sig.	Stat.	<i>df</i>	Sig.
Vortest	Schwierigkeit	KG	.117	141	.000	.979	141	.028
		VG	.126	122	.000	.977	122	.034
	Geschlecht	KG	.176	141	.000	.922	141	.000
		VG	.185	122	.000	.891	122	.000
	Selbstkonzept	KG	.090	141	.007	.959	141	.000
		VG	.105	122	.002	.959	141	.001
Nachtest	Schwierigkeit	KG	.123	146	.000	.969	146	.002
		VG	.100	133	.002	.980	133	.051
	Geschlecht	KG	.158	146	.000	.911	146	.000
		VG	.153	133	.000	.925	133	.000
	Selbstkonzept	KG	.086	146	.010	.974	146	.008
		VG	.065	133	.200	.980	133	.051

Anmerkung Signifikanzen ($p < .05$) sind fett hervorgehoben.

Tabelle C.7 Normalverteilungstests aller Interessesskalenmittelwerte im Vortest

		Kolmogorov-Smirnov			Shapiro-Wilk		
		Stat.	df	Sig.	Stat.	df	Sig.
wse	KG	.132	131	.000	.964	131	.002
	VG	.105	109	.005	.968	109	.010
wku	KG	.111	131	.000	.969	131	.004
	VG	.117	109	.001	.950	109	.000
wau	KG	.179	131	.000	.941	131	.000
	VG	.148	109	.000	.925	109	.000
wsewkuwau	KG	.059	142	.200	.982	142	.056
	VG	.091	121	.016	.975	121	.024
JOYSCIE	KG	.069	131	.200*	.975	131	.016
	VG	.093	109	.020	.936	109	.000
INSTSCIE	KG	.080	131	.039	.975	131	.016
	VG	.095	109	.017	.961	109	.003
INTSCIE	KG	.113	131	.000	.978	131	.030
	VG	.100	109	.010	.977	109	.053
mo1	KG	.122	131	.000	.947	131	.000
	VG	.160	109	.000	.897	109	.000
mo2	KG	.077	131	.054	.982	131	.077
	VG	.141	109	.000	.958	109	.002
mo3	KG	.103	131	.002	.965	131	.002
	VG	.122	109	.000	.937	109	.000
mo123	KG	.051	142	.200	.990	142	.393
	VG	.097	121	.007	.970	121	.009
meco	KG	.110	131	.000	.975	131	.014
	VG	.091	109	.026	.968	109	.010

Tabelle C.8 Normalverteilungstests aller Interessesskalenmittelwerte im Nachtest

		Kolmogorov-Smirnov			Shapiro-Wilk		
		Stat.	df	Sig.	Stat.	df	Sig.
wse	KG	.132	129	.000	.970	129	.006
	VG	.151	114	.000	.962	114	.003
wku	KG	.139	129	.000	.973	129	.011
	VG	.146	114	.000	.958	114	.001
wau	KG	.137	129	.000	.946	129	.000
	VG	.168	114	.000	.944	114	.000
wsewkuwau	KG	.076	129	.067	.984	129	.124
	VG	.107	114	.003	.971	114	.014
JOYSCIE	KG	.117	129	.000	.970	129	.006
	VG	.116	114	.001	.963	114	.003
INSTSCIE	KG	.118	129	.000	.964	129	.002
	VG	.095	114	.013	.973	114	.020
INTSCIE	KG	.085	129	.022	.988	129	.350
	VG	.076	114	.110	.963	114	.003
mo1	KG	.095	129	.006	.969	129	.005
	VG	.112	114	.001	.945	114	.000
mo2	KG	.104	129	.002	.983	129	.099
	VG	.091	114	.021	.980	114	.086
mo3	KG	.172	129	.000	.929	129	.000
	VG	.103	114	.004	.965	114	.004
mo123	KG	.080	129	.041	.983	129	.111
	VG	.081	114	.062	.984	114	.198
meco	KG	.095	129	.007	.977	129	.025
	VG	.112	114	.001	.978	114	.052

Tabelle C.9 Normalverteilungstests aller Fachwissen-Personenparamter im Vor- und Nachtest

		Kolmogorov-Smirnov			Shapiro-Wilk		
		Stat.	df	Sig.	Stat.	df	Sig.
FW Vortest	KG	.089	144	.007	.967	144	.001
	VG	.076	119	.089	.983	119	.134
FW Nachtest	KG	.116	144	.000	.959	144	.000
	VG	.131	119	.000	.959	119	.001

Anmerkung Signifikanzen ($p < .05$) sind fett hervorgehoben.

C.3 Tests auf Varianzhomogenität

Tabelle C.10 Levene-Test auf Gleichheit der Fehlervarianzen der drei IAT-Dimensionen

		<i>F</i>	<i>df</i> 1	<i>df</i> 2	Sig.
Schwierigkeit	Vortest	3.298	1	217	.071
	Nachtest	2.378	1	217	.124
Geschlechtskonnotation	Vortest	0.335	1	182	.564
	Nachtest	0.225	1	182	.635
Selbstkonzept	Vortest	0.190	1	195	.663
	Nachtest	1.193	1	195	.276

Tabelle C.11 Levene-Test auf Gleichheit der Fehlervarianzen der drei semantischen Differentiale

		<i>F</i>	<i>df</i> 1	<i>df</i> 2	Sig.
Schwierigkeit	Vortest	0.123	1	238	.726
	Nachtest	0.820	1	238	.366
Geschlechtskonnotation	Vortest	0.018	1	238	.892
	Nachtest	0.300	1	238	.585
Selbstkonzept	Vortest	0.528	1	238	.468
	Nachtest	0.008	1	238	.930

Tabelle C.12 Levene-Test auf Gleichheit der Fehlervarianzen im Fachwissenstest

		<i>F</i>	<i>df</i> 1	<i>df</i> 2	Sig.
Fachwissen	Vortest	0.867	1	259	.353
	Nachtest	0.032	1	259	.858

Tabelle C.13 Levene-Test auf Gleichheit der Fehlervarianzen der Interessesskalen

		<i>F</i>	<i>df</i> 1	<i>df</i> 2	Sig.
wse	Vortest	0.945	1	225	.332
	Nachtest	0.156	1	225	.693
wku	Vortest	0.744	1	225	.389
	Nachtest	0.369	1	225	.544
wau	Vortest	0.787	1	229	.376
	Nachtest	0.129	1	229	.720
wsewkuwau	Vortest	0.262	1	233	.609
	Nachtest	0.210	1	233	.647
JOYSCIE	Vortest	1.087	1	231	.298
	Nachtest	0.078	1	231	.780
INSTSCIE	Vortest	5.019	1	215	.026
	Nachtest	0.084	1	215	.772
INTSCIE	Vortest	0.013	1	231	.909
	Nachtest	0.003	1	231	.955
mo1	Vortest	0.053	1	231	.819
	Nachtest	0.143	1	231	.705
mo2	Vortest	1.395	1	229	.239
	Nachtest	0.233	1	229	.630
mo3	Vortest	9.010	1	230	.003
	Nachtest	0.366	1	230	.546
mo123	Vortest	1.181	1	232	.278
	Nachtest	0.806	1	232	.370
meco	Vortest	3.170	1	223	.076
	Nachtest	0.089	1	223	.766

Anmerkung Signifikanzen ($p < .05$) sind fett hervorgehoben.

C.4 Vorannahmen ANCOVA

Tabelle C.14 Vergleich der Vorteststatistiken zur Überprüfung der Unabhängigkeit der Kovariate vom Zwischensubjektfaktor

		<i>T</i>	<i>df</i>	Sig.	Diff.	SE Diff.
Implizit	Schwierigkeit	0.211	231.6	.833	0.009	0.041
	Geschlecht	-0.644	238	.520	-0.028	0.044
	Selbstkonzept	0.781	245	.436	0.029	0.038
Explizit	Schwierigkeit	-3.540	261	.000	-0.870	0.246
	Geschlecht	-1.329	261	.185	-0.283	0.213
	Selbstkonzept	-2.998	261	.003	-0.987	0.329
Interesse	wse	1.709	260	.089	0.165	0.097
	wku	2.763	258	.006	0.253	0.092
	wau	0.737	259	.462	0.063	0.086
	wsewkuwau	2.317	263	.021	0.169	0.073
	JOYSCIE	0.725	263	.469	0.068	0.094
	INSTSCIE	0.901	237.2	.368	0.092	0.102
	INTSCIE	0.390	260	.697	0.025	0.064
	mo1	1.866	260	.063	0.122	0.066
	mo2	1.540	260	.125	0.136	0.089
	mo3	2.349	231.7	.020	0.241	0.103
	mo123	2.449	261	.015	0.175	0.071
	meco	-0.335	223.0	.738	-0.032	0.096
Fachwissen		-2.154	261	.032	-0.343	0.159

Anmerkung Signifikanzen ($p < .05$) sind fett hervorgehoben.

Tabelle C.15 Überprüfung der Interaktion von Zwischensubjektfaktor (Gruppe) und Kovariate (Vor-teststatistik) in den Fällen, in denen beide Variablen unabhängig voneinander sind (siehe Tabelle C.14)

	QS Typ III	<i>df</i>	<i>F</i>	Sig.
Gruppe \times D_{Schw}	0.020	1	0.186	.667
Gruppe \times D_{Ge}	0.001	1	0.009	.926
Gruppe \times D_{Se}	0.083	1	0.962	.328
Gruppe \times Explizit Ge	0.057	1	0.023	.880
Gruppe \times wse	0.276	1	0.698	.404
Gruppe \times wau	0.021	1	0.057	.811
Gruppe \times JOYSCIE	0.281	1	1.300	.255
Gruppe \times INSTSCIE	0.580	1	1.544	.215
Gruppe \times INTSCIE	0.103	1	0.618	.432
Gruppe \times mo1	0.789	1	4.065	.045
Gruppe \times mo2	0.987	1	3.837	.051
Gruppe \times meco	0.004	1	0.013	.910

Anmerkung Signifikanzen ($p < .05$) sind fett hervorgehoben.

C.5 Deskriptive Statistiken

Tabelle C.16 Deskriptive Statistiken der *D*-Werte im Vor- und Nachtest

			N	SpW ^a	Min	Max	M	SE	SD
Vortest	KG	D_{Schw}	140	1.52	-0.60	0.93	0.18	0.02	0.30
		D_{Ge}	132	1.27	-0.61	0.66	0.12	0.03	0.31
		D_{Se}	133	1.61	-0.79	0.81	0.01	0.02	0.29
	VG	D_{Schw}	120	1.66	-0.70	0.97	0.18	0.03	0.36
		D_{Ge}	108	1.98	-1.01	0.97	0.09	0.04	0.37
		D_{Se}	114	1.50	-0.59	0.91	0.04	0.03	0.30
Nachtest	KG	D_{Schw}	127	1.52	-0.56	0.95	0.26	0.03	0.33
		D_{Ge}	114	1.71	-0.69	1.03	0.15	0.03	0.32
		D_{Se}	117	1.33	-0.62	0.71	0.05	0.03	0.28
	VG	D_{Schw}	123	1.87	-0.80	1.07	0.26	0.03	0.37
		D_{Ge}	101	1.31	-0.55	0.76	0.12	0.03	0.30
		D_{Se}	113	1.57	-0.74	0.83	0.06	0.03	0.32

Anmerkung ^aSpannweite

Tabelle C.17 Deskriptive Statistiken der Mittelwerte der semantischen Differentiale im Vor- und Nachtest

			<i>N</i>	SpW ^a	Min	Max	<i>M</i>	<i>SE</i>	<i>SD</i>
Vortest	KG	<i>M_{Schw}</i>	141	10.0	0.0	10.0	5.36	0.17	1.99
		<i>M_{Ge}</i>	141	7.5	2.5	10.0	6.44	0.14	1.66
		<i>M_{Se}</i>	141	10.0	0.0	10.0	5.19	0.23	2.73
	VG	<i>M_{Schw}</i>	122	10.0	0.0	10.0	4.49	0.18	1.99
		<i>M_{Ge}</i>	122	10.0	0.0	10.0	6.16	0.16	1.79
		<i>M_{Se}</i>	122	10.0	0.0	10.0	4.20	0.23	2.58
Nachtest	KG	<i>M_{Schw}</i>	146	10.0	0.0	10.0	5.28	0.19	2.29
		<i>M_{Ge}</i>	146	10.0	0.0	10.0	6.60	0.15	1.79
		<i>M_{Se}</i>	146	10.0	0.0	10.0	5.53	0.21	2.54
	VG	<i>M_{Schw}</i>	133	10.0	0.0	10.0	4.68	0.18	2.10
		<i>M_{Ge}</i>	133	10.0	0.0	10.0	6.28	0.17	1.91
		<i>M_{Se}</i>	133	10.0	0.0	10.0	4.83	0.21	2.47

Anmerkung ^aSpannweite**Tabelle C.18** Deskriptive Statistiken der Personenparameter (Fachwissen) im Vor- und Nachtest

		<i>N</i>	SpW ^a	Min	Max	<i>M</i>	<i>SE</i>	<i>SD</i>
Vortest	KG	144	6.15	-1.82	4.33	0.97	0.11	1.33
	VG	119	6.32	-1.99	4.33	0.62	0.11	1.23
Nachtest	KG	144	5.98	-1.66	4.32	1.58	0.11	1.33
	VG	119	7.38	-3.06	4.32	1.16	0.12	1.35

Anmerkung ^aSpannweite

Tabelle C.19 Deskriptive Statistiken der Mittelwerte der Interessesskalen im Vortest

		<i>N</i>	SpW ^a	Min	Max	<i>M</i>	<i>SE</i>	<i>SD</i>
Kontrollgruppe	wse	141	3.00	1.00	4.00	2.55	0.07	0.80
	wku	140	3.00	1.00	4.00	2.65	0.06	0.74
	wau	142	3.00	1.00	4.00	3.02	0.06	0.66
	wsewkuwau	142	2.89	1.00	3.89	2.74	0.05	0.57
	JOYSCIE	142	3.00	1.00	4.00	2.80	0.06	0.71
	INSTSCIE	135	3.00	1.00	4.00	2.56	0.06	0.74
	INTSCIE	141	2.50	1.50	4.00	2.92	0.04	0.51
	mo1	141	2.83	1.17	4.00	3.05	0.04	0.50
	mo2	142	2.83	1.17	4.00	2.68	0.06	0.68
	mo3	142	3.00	1.00	4.00	2.52	0.06	0.74
	mo123	142	2.80	1.20	4.00	2.77	0.05	0.56
	meco	142	3.00	1.00	4.00	2.72	0.06	0.69
Versuchsgruppe	wse	121	3.00	1.00	4.00	2.71	0.07	0.76
	wku	120	3.00	1.00	4.00	2.91	0.07	0.74
	wau	119	3.00	1.00	4.00	3.08	0.07	0.73
	wsewkuwau	123	2.44	1.56	4.00	2.91	0.06	0.62
	JOYSCIE	123	3.00	1.00	4.00	2.87	0.07	0.82
	INSTSCIE	122	3.00	1.00	4.00	2.65	0.08	0.88
	INTSCIE	121	2.50	1.50	4.00	2.95	0.05	0.53
	mo1	121	3.00	1.00	4.00	3.18	0.05	0.56
	mo2	120	3.00	1.00	4.00	2.81	0.07	0.75
	mo3	120	3.00	1.00	4.00	2.77	0.08	0.89
	mo123	121	2.82	1.18	4.00	2.94	0.05	0.60
	meco	116	3.00	1.00	4.00	2.69	0.08	0.83

Anmerkung ^aSpannweite

Tabelle C.20 Deskriptive Statistiken der Mittelwerte der Interessesskalen im Nachtest

		<i>N</i>	SpW ^a	Min	Max	<i>M</i>	<i>SE</i>	<i>SD</i>
Kontrollgruppe	wse	142	3.00	1.00	4.00	2.60	0.06	0.71
	wku	143	3.00	1.00	4.00	2.70	0.06	0.72
	wau	144	2.67	1.33	4.00	3.00	0.06	0.68
	wsewkuwau	146	2.71	1.17	3.88	2.77	0.05	0.56
	JOYSCIE	145	3.00	1.00	4.00	2.68	0.06	0.73
	INSTSCIE	137	3.00	1.00	4.00	2.43	0.07	0.80
	INTSCIE	144	2.71	1.29	4.00	2.87	0.04	0.53
	mo1	145	2.83	1.17	4.00	2.95	0.04	0.54
	mo2	145	3.00	1.00	4.00	2.53	0.06	0.69
	mo3	146	3.00	1.00	4.00	2.41	0.06	0.74
	mo123	146	2.81	1.19	4.00	2.65	0.05	0.57
	meco	144	2.75	1.25	4.00	2.67	0.06	0.70
Versuchsgruppe	wse	125	3.00	1.00	4.00	2.76	0.06	0.66
	wku	128	3.00	1.00	4.00	2.84	0.06	0.70
	wau	129	3.00	1.00	4.00	2.95	0.06	0.69
	wsewkuwau	129	2.64	1.25	3.89	2.84	0.05	0.54
	JOYSCIE	126	3.00	1.00	4.00	2.77	0.06	0.70
	INSTSCIE	123	3.00	1.00	4.00	2.50	0.07	0.81
	INTSCIE	127	2.73	1.13	3.86	2.87	0.05	0.52
	mo1	126	2.67	1.17	3.83	3.09	0.04	0.48
	mo2	124	3.00	1.00	4.00	2.59	0.06	0.67
	mo3	125	3.00	1.00	4.00	2.57	0.07	0.75
	mo123	126	2.18	1.59	3.76	2.77	0.04	0.47
	meco	122	3.00	1.00	4.00	2.70	0.06	0.67

Anmerkung ^aSpannweite

C.6 Varianzanalysen

Tabelle C.21 Zweifaktorielle Varianzanalyse der *D*-Werte

		<i>df</i>	<i>F</i>	Sig.	η_p^2
Schwierigkeit	Testzeit	1	7.695	.006	.034
	Gruppe	1	0.050	.824	.000
	Testzeit \times Gruppe	1	0.176	.676	.001
	Fehler (Testzeit)	217			
Geschlechtskonnotation	Testzeit	1	0.651	.421	.004
	Gruppe	1	0.028	.867	.000
	Testzeit \times Gruppe	1	0.012	.913	.000
	Fehler (Testzeit)	182			
Selbstkonzept	Testzeit	1	1.212	.272	.006
	Gruppe	1	1.102	.295	.006
	Testzeit \times Gruppe	1	1.472	.227	.007
	Fehler (Testzeit)	195			

Anmerkung Signifikanzen ($p < .05$) sind fett hervorgehoben.

Tabelle C.22 Zweifaktorielle Varianzanalyse der semantischen Differentiale

		<i>df</i>	<i>F</i>	Sig.	η_p^2
Schwierigkeit	Testzeit	1	0.184	.668	.001
	Gruppe	1	11.772	.001	.047
	Testzeit \times Gruppe	1	2.341	.127	.010
	Fehler (Testzeit)	238			
Geschlechtskonnotation	Testzeit	1	1.447	.230	.006
	Gruppe	1	1.559	.213	.007
	Testzeit \times Gruppe	1	0.497	.568	.001
	Fehler (Testzeit)	238			
Selbstkonzept	Testzeit	1	15.250	.000	.060
	Gruppe	1	8.281	.004	.034
	Testzeit \times Gruppe	1	2.806	.191	.007
	Fehler (Testzeit)	238			

Anmerkung Signifikanzen ($p < .05$) sind fett hervorgehoben.

Tabelle C.23 Zweifaktorielle Varianzanalyse der Interessesskalen (Teil 1)

		<i>df</i>	<i>F</i>	Sig.	η_p^2
Wahrgenommene Soziale Eingebundenheit	Testzeit	1	0.581	.447	.003
	Gruppe	1	5.367	.021	.023
	Testzeit \times Gruppe	1	1.176	.279	.005
	Fehler (Testzeit)	225			
Wahrgenommene Kompetenzunterstützung	Testzeit	1	0.591	.443	.003
	Gruppe	1	8.157	.005	.035
	Testzeit \times Gruppe	1	2.233	.136	.010
	Fehler (Testzeit)	225			
Wahrgenommene Autonomieunterstützung	Testzeit	1	3.813	.052	.016
	Gruppe	1	0.020	.888	.000
	Testzeit \times Gruppe	1	2.416	.122	.010
	Fehler (Testzeit)	229			
Unterstützende Lehr-/Lernbedingungen	Testzeit	1	1.287	.258	.005
	Gruppe	1	4.046	.045	.017
	Testzeit \times Gruppe	1	3.969	.048	.017
	Fehler (Testzeit)	233			
Freude und Interesse an Naturwissenschaften	Testzeit	1	15.044	.000	.061
	Gruppe	1	0.995	.320	.004
	Testzeit \times Gruppe	1	0.235	.628	.001
	Fehler (Testzeit)	231			
Instrumentelle physikbezogene Motivation	Testzeit	1	17.204	.000	.074
	Gruppe	1	0.000	.992	.000
	Testzeit \times Gruppe	1	0.767	.382	.004
	Fehler (Testzeit)	215			
Interesse über Naturwissenschaften zu lernen	Testzeit	1	3.718	.055	.016
	Gruppe	1	0.008	.929	.000
	Testzeit \times Gruppe	1	1.891	.170	.008
	Fehler (Testzeit)	231			

Anmerkung Signifikanzen ($p < .05$) sind fett hervorgehoben.

Tabelle C.24 Zweifaktorielle Varianzanalyse der Interessesskalen (Teil 2)

		<i>df</i>	<i>F</i>	Sig.	η_p^2
Amotiviertheit/ Externale Motiviertheit	Testzeit	1	9.212	.003	.038
	Gruppe	1	5.857	.016	.025
	Testzeit \times Gruppe	1	0.722	.396	.003
	Fehler (Testzeit)	231			
Introjizierte/ Identifizierte Motiviertheit	Testzeit	1	24.243	.000	.096
	Gruppe	1	1.368	.243	.006
	Testzeit \times Gruppe	1	2.434	.120	.011
	Fehler (Testzeit)	229			
Intrinsische/ Interessierte Motiviertheit	Testzeit	1	22.823	.000	.090
	Gruppe	1	3.039	.083	.013
	Testzeit \times Gruppe	1	5.371	.021	.023
	Fehler (Testzeit)	230			
Qualität der Lernmotivation	Testzeit	1	33.439	.000	.126
	Gruppe	1	4.628	.032	.020
	Testzeit \times Gruppe	1	5.034	.026	.021
	Fehler (Testzeit)	232			
Wahrgenommene Beudeutung der Lerninhalte	Testzeit	1	0.256	.614	.001
	Gruppe	1	0.005	.943	.000
	Testzeit \times Gruppe	1	0.275	.600	.001
	Fehler (Testzeit)	223			

Anmerkung Signifikanzen ($p < .05$) sind fett hervorgehoben.

Tabelle C.25 Zweifaktorielle Varianzanalyse der Personenparameter (Fachwissen)

	<i>df</i>	<i>F</i>	Sig.	η_p^2
Testzeit	1	46.085	.000	.151
Gruppe	1	7.615	.006	.029
Testzeit \times Gruppe	1	0.154	.695	.002
Fehler (Testzeit)	259			

Anmerkung Signifikanzen ($p < .05$) sind fett hervorgehoben.

C.7 Kovarianzanalysen

Tabelle C.26 Kovarianzanalyse der *D*-Werte (Nachtest) mit den Vortestergebnissen als Kovariate

		<i>df</i>	<i>F</i>	Sig.	η_p^2
Schwierigkeit	Vortest	1	32.839	.000	.133
	Gruppe	1	0.169	.681	.001
	Fehler	215			
Geschlechtskonnotation	Vortest	1	5.802	.017	.031
	Gruppe	1	0.026	.873	.000
	Fehler	180			
Selbstkonzept	Vortest	1	4.802	.030	.024
	Gruppe	1	0.084	.772	.000
	Fehler	193			

Tabelle C.27 Kovarianzanalyse der Mittelwerte der semantischen Differentiale (Nachtest) mit den Vortestergebnissen als Kovariate

		<i>df</i>	<i>F</i>	Sig.	η_p^2
Geschlechtskonnotation	Vortest	1	92.766	.000	.281
	Gruppe	1	0.002	.962	.000
	Fehler	237			

Tabelle C.28 Kovarianzanalyse der Mittelwerte der Interessesskalen (Nachtest) mit den Vortestergebnissen als Kovariate

		<i>df</i>	<i>F</i>	Sig.	η_p^2
wse	Vortest	1	47.313	.000	.174
	Gruppe	1	0.196	.658	.001
	Fehler	224			
wau	Vortest	1	45.974	.000	.168
	Gruppe	1	1.485	.224	.006
	Fehler	228			
JOYSCIE	Vortest	1	317.706	.000	.580
	Gruppe	1	0.005	.944	.000
	Fehler	230			
INSTSCIE	Vortest	1	143.299	.000	.401
	Gruppe	1	0.631	.428	.003
	Fehler	214			
INTSCIE	Vortest	1	167.378	.000	.421
	Gruppe	1	1.625	.204	.007
	Fehler	230			
mo2	Vortest	1	152.783	.000	.401
	Gruppe	1	0.699	.404	.003
	Fehler	228			
meco	Vortest	1	82.066	.000	.270
	Gruppe	1	0.179	.673	.001
	Fehler	222			

C.8 Post-Hoc-Teststärkeanalysen

Tabelle C.29 Post-Hoc-Teststärkeanalysen aller untersuchten Statistiken für die Interaktion einer zweifaktoriellen ANOVA mit Messwiederholung

		Power ^a	beob. Effekt	beob. Power ^b	Sensitiv ^c für η_p^2	
Implizit	Schwierigkeit	.99	.001	.07	.02	mittlere Effekte
	Selbstkonzept	.99	.007	.23	.03	
	Geschlecht	.99	.000	.05	.03	
Explizit	Schwierigkeit	.99	.010	.33	.01	schwache Effekte
	Selbstkonzept	.99	.007	.26	.01	
	Geschlecht	.99	.001	.09	.01	
Interesse	wse	.99	.005	.19	.02	mittlere Effekte
	wku	.99	.010	.32	.03	
	wau	.99	.010	.34	.03	
	wsewkuwau	.99	.017	.51	.03	
	JOYSCIE	.99	.001	.08	.03	
	INSTSCIE	.99	.004	.14	.03	
	INTSCIE	.99	.008	.28	.03	
	mo1	.99	.003	.14	.03	
	mo2	.99	.011	.34	.03	
	mo3	.99	.023	.64	.03	
	mo123	.99	.021	.61	.03	
	meco	.99	.001	.08	.03	
Fachwissen		.99	.001	.08	.02	

Anmerkung ^afür mittlere Effekte, $\alpha = .05$

^bfür beobachtete Effektstärken

^cbei $\alpha = .05$ und $\beta = .05$

Abbildungsverzeichnis

2.1	Merkmalsaspekte von PbPU	11
2.2	Bedeutende Tätigkeiten im PbPU	15
2.3	INVO-Modell	27
2.4	Themenübersicht der phänomenbasierten Anfangsoptik	33
2.5	Der weiße Schatten	37
3.1	Geschlechtsbezogene Unterschiede im Interesse an Physik	45
3.2	Geschlechtsspezifisches Interesse an den MINT-Fächern	46
3.3	Männlichkeit und Weiblichkeit von Physik-Prototypen	49
3.4	Assoziationen zu Physik	51
4.1	Verteilung der Latenzzeiten eines IATs	59
4.2	Ablauf eines IATs mit sieben Blöcken	60
4.3	Zwei beispielhafte Trials eines ST-IATs	74
5.1	Eine <i>Social Knowledge Structure</i> (SKS)	80
5.2	Darstellung des Balanced Identity Designs (BID)	84
5.3	Mehrdimensionales Selbstkonzept	85
5.4	Ausgangslage und Ziel für das Image von Physik	88
7.1	Geplantes Forschungsdesign	98
8.1	Flash-Simulation zur abgelösten und eingebundenen Perspektive	111
8.2	Experimenteller Aufbau zur Spiegelwelt	116
8.3	Tafelbild zur Herleitung des 2. Spiegelgesetzes	118
9.1	Design der zweiten Pilotstudie	126

11.1	Interaktionsdiagramme der drei IAT-Dimensionen	141
11.2	Interaktionsdiagramme aller IAT-Dimensionen	142
11.3	Interaktionsdiagramme der expliziten Assoziationen	144
11.4	Vergleich der Polaritätsprofile	145
11.5	Interaktionsdiagramme der Skala <i>wsewkuwau</i>	146
11.6	Interaktionsdiagramme der Skalen <i>JOYSCIE</i> , <i>INSTSCIE</i> , <i>INTSCIE</i> und <i>meco</i>	147
11.7	Interaktionsdiagramme der Skala <i>mo123</i>	148
11.8	Interaktionsdiagramme der Personenparameter	150
11.9	Interaktionsdiagramme (Testzeit \times Lehrer)	151
11.10	Geschlechtsdifferenziertes Polaritätsprofil	152
11.11	Interaktionsdiagramm (Testzeit \times Geschlecht \times Gruppe)	153
C.1	Split-Half-Analyse der Itemparameter	213
C.2	Überdeckung der Personen- und Itemparameterverteilungen	213

Tabellenverzeichnis

4.1	Datenaufbereitung und Berechnung des D-Wertes	66
7.1	Übersicht über die eingesetzten Interessesskalen mit Beispielitems.	104
10.1	Zusammensetzung der Stichprobe	130
A.1	Übersicht der Stimuli in den drei eingesetzten IATs	171
C.1	Verwendete Effektstärkemaße	211
C.2	Cronbachs α -Koeffizienten für alle IATs	212
C.3	Cronbachs α -Koeffizienten für alle Interessesskalen	212
C.4	Itemkennwerte der Raschanalyse	212
C.5	Normalverteilungstests der <i>D</i> -Werte	214
C.6	Normalverteilungstests der semantischen Differentiale	214
C.7	Normalverteilungstests der Interessesskalen (Vortest)	215
C.8	Normalverteilungstests der Interessesskalen (Nachtest)	216
C.9	Normalverteilungstests der Fachwissen-Personenparameter	216
C.10	Levene-Test IAT-Dimensionen	217
C.11	Levene-Test semantischen Differentiale	217
C.12	Levene-Test Fachwissenstest	217
C.13	Levene-Test Interessesskalen	218
C.14	Überprüfung der Unabhängigkeit der Kovariate	219
C.15	Überprüfung der Interaktion von Zwischensubjektfaktor	220
C.16	Deskriptive Statistiken der <i>D</i> -Werte	220
C.17	Deskriptive Statistiken der semantischen Differentiale	221
C.18	Deskriptive Statistiken der Personenparameter (Fachwissen)	221
C.19	Deskriptive Statistiken der Interessesskalen (Vortest)	222

C.20 Deskriptive Statistiken der Interessesskalen (Nachtest)	223
C.21 Zweifaktorielle Varianzanalyse der <i>D</i> -Werte	224
C.22 Zweifaktorielle Varianzanalyse der semantischen Differentiale	224
C.23 Zweifaktorielle Varianzanalyse der Interessesskalen (Teil 1)	225
C.24 Zweifaktorielle Varianzanalyse der Interessesskalen (Teil 2)	226
C.25 Zweifaktorielle Varianzanalyse der Personenparameter (Fachwissen) . .	226
C.26 Kovarianzanalyse der <i>D</i> -Werte	227
C.27 Kovarianzanalyse der semantischen Differentiale	227
C.28 Kovarianzanalyse der Interessesskalen	228
C.29 Post-Hoc-Teststärkeanalysen	229

- Aufschnaiter, C. v. (2005). Von Fehlvorstellungen und fehlenden Erfahrungen: Konzeptentwicklung im Physik-Unterricht. In V. Grötzebauch & H. Nordmeier (Hrsg.), *Didaktik der Physik: Berlin 2005*. Deutsche Physikalische Gesellschaft. DPG-Frühjahrstagung. Berlin: Lehmanns Media.
- Barz, H. & Randoll, D. (Hrsg.). (2007). *Absolventen von Waldorfschulen: Eine empirische Studie zu Bildung und Lebensgestaltung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Bluemke, M. & Friese, M. (2006). Do features of stimuli influence IAT effects? *Journal of Experimental Social Psychology*, 42, 163–176.
- Bluemke, M. & Friese, M. (2008). Reliability and validity of the Single-Target IAT (ST-IAT): Assessing automatic affect towards multiple attitude objects. *European Journal of Social Psychology*, 38, 977–997.
- Bockemühl, M. (1995). Güterwagen-Schieben: Grunderfahrungen vom Zusammenhang zwischen Kraft und Beschleunigung. *Naturwissenschaften im Unterricht Physik*, 6(29) Physik erleben, 134–135.
- Bond, T. G. & Fox, C. M. (2007). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences* (2. Auflage). New York, London: Routledge.
- Bortz, J. & Döring, N. (2002). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (3. Auflage). Berlin; Heidelberg: Springer.
- Bosson, J. K., Swann Jr., W. B. & Penebaker, J. W. (2000). Stalking the perfect measure of Implicit Self-Esteem: The blind men and the elephant revisited? *Journal of Personality and Social Psychology*, 79(4), 631–643.
- Buck, P. & Mackensen, M. v. (2006). *Naturphänomene erlebend verstehen* (7. Auflage). Köln: Aulis Verlag Deubner.
- Cohen, J. (1992). A Power Primer. *Psychological Bulletin*, 112(1), 155–159.

- Cornelsen Schulverlage GmbH. (2014). Flash Simulation Licht und Schatten 2. Zugriff 4. März 2014, unter <http://www.cornelsen.de/interaktiv/1.c.891157.de?hasjs=1393947581&submittedByForm=1>
- Cvencek, D., Greenwald, A. G. & Meltzoff, A. N. (2012). Balanced Identity Theory: Review of evidence for implicit consistency in social cognition. In B. Gawronski & F. Strack (Hrsg.), *Cognitive Consistency: A Fundamental Principle in Social Cognition* (Kap. 8, S. 157–177). The Guilford Press.
- Cvencek, D., Meltzoff, A. N. & Greenwald, A. G. (2011). Math–Gender stereotypes in elementary school children. *Child Development*, 82(3), 766–779.
- Deci, E. L. & Ryan, R. M. (1985). *Intrinsic Motivation and Self-Determination in Human Behavior*. New York: Plenum Press.
- Deci, E. L. & Ryan, R. M. (1993). Die Selbstbestimmungstheorie der Motivation und ihre Bedeutung für die Pädagogik. *Zeitschrift für Pädagogik*, 39(2), 223–238.
- Dimitrov, D. M. & Rumrill, P. D., Jr. (2003). Pretest-posttest designs and measurement of change. *Work*, 20, 159–165.
- Dotsch, R. & Wigboldus, D. H. J. (2008). Virtual prejudice. *Journal of Experimental Social Psychology*, 44, 1194–1198.
- Dugard, P. & Todman, J. (1995). Analysis of Pre-test–Post-test Control Group Designs in Educational Research. *Educational Psychology*, 15(2), 181–198.
- Dunham, Y., Baron, A. S. & Banaji, M. R. (2006). From American city to Japanese village: A cross-cultural investigation of implicit race attitudes. *Child Development*, 77, 1268–1281.
- Eid, M., Gollwitzer, M. & Schmitt, M. (2011). *Statistik und Forschungsmethoden* (2. korrigierte Auflage). Weinheim, Basel: Beltz Verlag.
- Engelbrecht, A. (2003). *Kritik der Pädagogik Martin Wagenscheins: Eine Reflexion seines Beitrages zur Didaktik*. Didaktik. Münster: LIT Verlag.
- Erb, R. (1994). *Optik mit Lichtwegen – das Fermat-Prinzip als Grundlage für das Verstehen der Optik*. Bochum: Westarp-Wissenschaften.
- Erb, R. & Schön, L. (1996). Vom Sehen zur Optik: Ein Curriculum für die Mittelstufe. *Praxis der Naturwissenschaften – Physik*, 45(8), 31–36.
- Faul, F., Erdfelder, E., Lang, A.-G. & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175–191.
- Fazio, R. H. (2001). On the automatic activation of associated evaluations: An overview. *Cognition and Emotion*, 15(2), 115–141.
- Fazio, R. H. & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology*, 54, 297–327.
- Fedra, D. & Schön, L. (1989). Sinneswahrnehmungen und Physik. *Physik und Didaktik*, 2, 104–123.

- Festinger, L. (1957). *A theory of cognitive dissonance*. Palo Alto, CA: Stanford University Press.
- Field, A. (2009). *Discovering Statistics Using SPSS* (3. Auflage). London; Thousand Oaks; New Delhi: SAGE Publications.
- Fischer, H. E., Schecker, H. & Wiesner, H. (2004). Kerncurriculum Physik. *Der mathematische und naturwissenschaftliche Unterricht*, 57(3), 147–154.
- Frenzel, A. C., Götz, T. & Pekrun, R. (2009). Emotionen. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (Kap. 9, S. 204–231). Heidelberg: Springer Medizin Verlag.
- Frey, A., Taskinen, P., Schütte, K., Prenzel, M., Artelt, C., Baumert, J., ... Pekrun, R. (Hrsg.). (2009). *Pisa 2006 Skalenhandbuch – Dokumentation der Erhebungsinstrumente*. Waxmann.
- Gawronski, B. & Conrey, F. R. (2004). Der Implizite Assoziationstest als Maß automatisch aktivierter Assoziationen: Reichweite und Grenzen. *Psychologische Rundschau*, 55(3), 118–126.
- Glass, G. V., Peckham, P. D. & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analysis of variance and covariance. *Review of Educational Research*, 42, 237–288.
- Google Inc. (2013). Autorenprofil von Anthony G. Greenwald auf Google Scholar. Zugriff 12. August 2013, unter <http://scholar.google.com/citations?user=U24uY0AAAAAJ&hl=en>
- Grebe-Ellis, J. (2005). *Grundzüge einer Phänomenologie der Polarisation* (L.-H. Schön & J. Grebe-Ellis, Hrsg.). Phänomenologie in der Naturwissenschaft. Berlin: Logos Verlag.
- Grebe-Ellis, J. (2006a). Anmerkungen zum Selbstverständnis der "Modellfreien Optik". In H. Grötzebach & V. Nordmeier (Hrsg.), *Didaktik der Physik: Kassel 2006*. Deutsche Physikalische Gesellschaft. DPG-Frühjahrstagung. Berlin: Lehmanns Media.
- Grebe-Ellis, J. (2006b). Ein Nachtrag: Von der Optik im Tastraum zu einer Optik des Sehens. In J. Grebe-Ellis & F. Theilmann (Hrsg.), *open eyes 2005: Ansätze und Perspektiven der phänomenologischen Optik* (Bd. 1, S. 235–245). Phänomenologie in der Naturwissenschaft. Berlin: Logos Verlag.
- Grebe-Ellis, J. (2007). Lesen im Buch der Natur – Zur Entwicklung einer phänomenologischen Lesekompetenz. In H. Grötzebach & V. Nordmeier (Hrsg.), *Didaktik der Physik: Regensburg 2007*. Deutsche Physikalische Gesellschaft. DPG-Frühjahrstagung. Berlin: Lehmanns Media.
- Grebe-Ellis, J., Theilmann, F. & Rang, M. (2009). Lichtspuren im Wasser – Ein Experiment zum Verhältnis von Brechung und Hebung. *Physik und Didaktik in Schule und Hochschule*, 8(3), 86–91.

- Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A. & Mellott, D. S. (2002). A Unified Theory of Implicit Attitudes, Stereotypes, Self-Esteem, and self-Concept. *Psychological Review*, 109(1), 3–25.
- Greenwald, A. G., McGhee, D. E. & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, 74, 1464–1481.
- Greenwald, A. G. (2012). Studies Showing Uses of the IAT with “Real-World” Subject Populations. Zugriff 12. August 2013, unter http://faculty.washington.edu/agg/pdf/Real-world_samples.pdf
- Greenwald, A. G., Banaji, M. R., Poehlmann, T. A. & Uhlmann, E. L. (2009). Understanding and Using the Implicit Association Test: III. Meta-Analysis of Predictive Validity. *Journal of Personality and Social Psychology*, 97(1), 17–41.
- Greenwald, A. G., Nosek, B. A. & Banaji, M. R. (2003). Understanding and Using the Implicit Association Test: I. An Improved Scoring Algorithm. *Journal of Personality and Social Psychology*, 85(2), 197–216.
- Hannover, B. & Kessels, U. (2002). Challenge the science-stereotype! Der Einfluss von Technik-Freizeitkursen auf das Naturwissenschaften-Stereotyp von Schülerinnen und Schülern. In M. Prenzel & J. Doll (Hrsg.), *Bildungsqualität von Schule: Schulische und außerschulische Bedingungen mathematischer, naturwissenschaftlicher und überfachlicher Kompetenzen* (Bd. 45, S. 341–358). Weinheim: Beltz
- Hannover, B. & Kessels, U. (2004). Self-to-prototype matching as a strategy for making academic choices. Why high school students do not like math and science. *Learning and Instruction*, 14(1), 51–67.
- Hasselhorn, M. & Gold, A. (2013). *Pädagogische Psychologie: Erfolgreiches Lernen und Lehren* (3. Auflage) (M. Hasselhorn, H. Heuer & S. Schneider, Hrsg.). Standards Psychologie. Stuttgart: Verlag W. Kohlhammer.
- Häußler, P., Hoffmann, L., Langeheine, R., Rost, J. & Sievers, K. (1996). Qualitative Unterschiede im Interesse an Physik und Konsequenzen für den Physikunterricht. *Zeitschrift für Didaktik der Naturwissenschaften*, 2(3), 57–69.
- Heider, F. (1958). *The psychology of interpersonal relations*. (deutsch: *Psychologie der interpersonalen Beziehungen*. Stuttgart: Klett, 1977). New York: Wiley.
- Heinzerling, H. (1995). Vom Sehen zu Optik: Ein Erfahrungsbericht der ersten Einheiten. *Naturwissenschaften im Unterricht Physik*, 6(29) Physik erleben, 11–15.
- Hoffmann, L., Häußler, P. & Lehrke, M. (1998). *Die IPN-Interessenstudie Physik*. Kiel: Institut für die Pädagogik der Naturwissenschaften.
- Hofmann, W., Gawronski, B., Gschwendner, T., Le, H. & Schmitt, M. (2005). A Meta-Analysis on the Correlation Between the Implicit Association Test and Explicit Self-Report Measures. *Personality and Social Psychology Bulletin*, 31(10), 1369–1385.

- Holland, R. W., Vries, M. d., Hermesen, B. & Knippenberg, A. v. (2012). Mood and the Attitude–Behavior link: The happy act on impulse, the sad think twice. *Social Psychological and Personality Science*, 3(3), 356–364. eprint: <http://spp.sagepub.com/content/3/3/356.full.pdf+html>
- Inquisit 3.0.4.0 [Computer Software]. (2010). Seattle, WA: Millisecond Software.
- Jansen, M., Schröders, U. & Stanat, P. (2013). Motivationale Schülermerkmale in Mathematik und den Naturwissenschaften. In H. A. Pant, P. Stanat, U. Schroeders, A. Roppelt, T. Siegle & C. Pöhlmann (Hrsg.), *IQB-Ländervergleich 2012: Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I* (Kap. 11). Münster: Waxmann. Zugriff unter <http://www.iqb.hu-berlin.de/laendervergleich/laendervergleich/lv2012/Bericht.pdf>
- Jelinek, D. & Sun, L.-L. (2003). *Does Waldorf Offer a Viable Form of Science Education: A Research Monograph*. California State University, College of Education. Sacramento, CA.
- Julius, F. H. (1984). *Entwurf einer Optik: Zur Phänomenologie des Lichts*. Menschenkunde und Erziehung. Verlag Freies Geistesleben.
- Karpinski, A. & Steinman, R. B. (2006). The Single Category Implicit Association Test as a Measure of Implicit Social Cognition. *Journal of Personality and Social Psychology*, 91(1), 16–32.
- Kauertz, A. (2008). *Schwierigkeitserzeugende Merkmale physikalischer Leistungstestaufgaben* (H. Niedderer & H. Fischler, Hrsg.). Studien zum Physiklernen. Berlin: Logos Verlag.
- Kessels, U. (2005). Fitting into the stereotype: How gender-stereotyped perceptions of prototypic peers relate to liking for school subjects. *European Journal of Psychology of Education*, 20(3), 309–323.
- Kessels, U. (2008). Physikinteresse gilt als unweiblich. Erklärung für die Unterrepräsentanz von Mädchen und Frauen im MINT-Bereich. *Schule im Blickpunkt*, (2), 24–27.
- Kessels, U. & Hannover, B. (2004a). Empfundene “Selbstnähe“ als Mediator zwischen Fähigkeitsselbstkonzept und Leistungskurswahlintentionen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 36(3), 130–138.
- Kessels, U. & Hannover, B. (2004b). Entwicklung schulischer Interessen als Identitätsregulation. In J. Doll & M. Prenzel (Hrsg.), *Bildungsqualität von Schule: Lehrerprofessionalisierung, Unterrichtsentwicklung und Schülerförderung als Strategien der Qualitätsverbesserung* (S. 398–412). Münster: Waxmann.
- Kessels, U. & Hannover, B. (2006). Zum Einfluss des Image von mathematisch-naturwissenschaftlichen Schulfächern auf die schulische Interessen- und Leistungsentwicklung. In M. Prenzel & L. Allolio-Näcke (Hrsg.), *Untersuchungen zur Bildungsqualität von Schule – Abschlussbericht des DFG-Schwerpunktprogramms* (S. 350–369). Münster: Waxmann.

- Kessels, U., Rau, M. & Hannover, B. (2006). What goes well with physics? Measuring and altering the image of science. *British Journal of Educational Psychology*, 74(4), 761–780.
- Kohl, K. (2013). Wagenschein-Archiv 2.0. Zugriff 15. Januar 2014, unter www.martin-wagenschein.de
- Köhnlein, W. (Hrsg.). (1998). *Der Vorrang des Verstehens: Beiträge zur Pädagogik Martin Wagenscheins*. Bad Heilbrunn: Klinkhardt.
- Krapp, A. (1992a). Das Interessenkonstrukt: Bestimmungsmerkmale der Interessenshandlung und des individuellen Interesses aus Sicht einer Person-Gegenstands-Konzeption. In A. Krapp & M. Prenzel (Hrsg.), *Interesse, Lernen, Leistung: Neuere Ansätze der pädagogisch-psychologischen Interessenforschung* (Kap. 13, S. 297–329). Münster: Aschendorff.
- Krapp, A. (1992b). Konzepte und Forschungsansätze zur Analyse des Zusammenhangs von Interesse, Lernen und Leistung. In A. Krapp & M. Prenzel (Hrsg.), *Interesse, Lernen, Leistung: Neuere Ansätze der pädagogisch-psychologischen Interessenforschung* (S. 9–52). Münster: Aschendorff.
- Krapp, A. (1998). Entwicklung und Förderung von Interessen im Unterricht. *Psychologie in Erziehung und Unterricht*, 44, 185–201.
- Krapp, A. (1999). Intrinsischen Lernmotivation und Interesse. Forschungsansätze und konzeptuelle Überlegungen. *Zeitschrift für Pädagogik*, 45(3), 387–406.
- Krapp, A. (2004). An educational-psychological theory of interest and its relation to SDT. In E. L. Deci & R. M. Ryan (Hrsg.), *Handbook of Self-Determination Research* (Kap. 18, S. 405–426). University Rochester Press.
- Krapp, A. (2005). Basic needs and the development of interest and intrinsic motivational orientations. *Learning and Instruction*, 15(5), 381–395.
- Krüger, G. (2011a). Die Spiegelwelt im phänomenorientierten Unterricht. Induktive Erkenntnisgewinnung zum Thema Spiegel. *Naturwissenschaften im Unterricht. Physik*, 22(123/124), 59–69.
- Krüger, G. (2011b). Wo sehen wir die Dinge? In R. Erb & J. Grebe-Ellis (Hrsg.), *Alles, worin der Mensch sich ernstlich einlässt, ist ein Unendliches: Physikdidaktische Miniaturen* [Lutz-Helmut Schön zu Ehren](Bd. 6, S. 69–83). Phänomenologie in der Naturwissenschaft. Berlin: Logos Verlag.
- Kuhn, J. (2014). Mehrebenenanalysen am Beispiel der Lernwirkung von Aufgaben. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (Kap. 24, S. 297–310). Berlin, Heidelberg: Springer-Verlag.
- Lane, K. A., Banaji, M. R., Nosek, B. A. & Greenwald, A. G. (2007). Understanding and Using the Implicit Association Test: IV. What We Know (So Far) about the

- Mehod. In B. Wittenbrink & N. Schwarz (Hrsg.), *Implicit Measures of Attitudes* (Kap. 3, S. 59–101). The Guilford Press.
- Leisner-Bodenthin, A. (2006). Zur Entwicklung von Modellkompetenz im Physikunterricht. *Zeitschrift für Didaktik der Naturwissenschaften*, 12, 91–109.
- Lienert, G. A. & Raatz, U. (1998). *Testaufbau und Testanalyse* (6. Auflage). Weinheim: Beltz.
- Linacre, J. M. (2014). Winsteps Help for Rasch Analysis. Zugriff 13. Februar 2014, unter <http://www.winsteps.com/winman/>
- Mackensen, M. v. (1992). *Klang, Helligkeit und Wärme: Phänomenologischer Physikunterricht, entwickelt aus Praxis und Theorie der Waldorfschule*. Kassel: Edition Waldorf.
- Mackensen, M. v. & Ohlendorf, H.-C. (1998). *Modellfreie Optik*. Kassel: Pädagogische Forschungsstelle.
- Maier, G. (2003). *Optik der Bilder* (5. Auflage). Verlag der Kooperative Dürnaue.
- Maier, G. (2004). *blicken-sehen-schauen: Beiträge zur Physik als Erscheinungswissenschaft* (1. Auflage) (J. Grebe-Ellis, Hrsg.). Dürnaue: Verlag der Kooperative Dürnaue.
- Marsh, H. W. (1990). A multidimensional, hierarchical model of self-concept: Theoretical and empirical justification. *Educational Psychology Review*, 2(2), 77–172.
- Martin, M. O., Mullis, I. V. & Chrostowsk, S. J. (Hrsg.). (2004). *TIMSS 2003 Technical Report*. TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.
- Mierke, J. (2004). *Kognitive Prozesse bei der indirekten Messung von Einstellungen mit dem Impliziten Assoziationstest* (Diss., Albert-Ludwigs-Universität Freiburg).
- Mikelskis, H. F., Schön, L.-H. & Wilke, H.-J. (Hrsg.). (2006). *Physik plus: Klassen 7/8 Berlin*. Berlin: Cornelsen Verlag.
- Miller, G. A. & Chapman, J. P. (2001). Misunderstanding Analysis of Covariance. *Journal of Abnormal Psychology*, 110(1), 40–48.
- Millisecond Software. (2013). Inquisit Task Library. Zugriff 8. August 2013, unter <http://www.millisecond.com/download/library/IAT/>
- Mitchell, J. P., Nosek, B. A. & Banaji, M. R. (2003). Contextual variations in implicit evaluation. *Journal of Experimental Psychology: General*, 132(3), 455–469.
- Möller, J. & Trautwein, U. (2009). Selbstkonzept. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (Kap. 8, S. 178–203). Heidelberg: Springer Medizin Verlag.
- Muckenfuß, H. (1995). *Lernen im sinnstiftenden Kontext*. Berlin: Cornelsen Verlag.
- Müller, M. & Schön, L.-H. (2009). „Eingebundene“ vs. „abgelöste“ Perspektive – Vorschlag einer Begriffsklärung. In D. Höttercke (Hrsg.), *Chemie- und Physikdidaktik*

- für die Lehramtsausbildung* (Bd. 29, S. 217–219). Gesellschaft für Didaktik der Chemie und Physik. Jahrestagung in Schwäbisch Gmünd 2008. Münster: LIT Verlag.
- Murmann, L. (2006). Phänomenologische Erkenntnistheorie und Fachdidaktik – Eine Verortung. In J. Grebe-Ellis & F. Theilmann (Hrsg.), *open eyes 2005: Ansätze und Perspektiven der phänomenologischen Optik* (Bd. 1, S. 111–125). Phänomenologie in der Naturwissenschaft. Berlin: Logos Verlag.
- Nosek, B. A. (2007). Implicit-Explicit Relations. *Current Directions in Psychological Science*, 16(2), 65–69.
- Nosek, B. A. (2005). Moderators of the Relationship Between Implicit and Explicit Evaluation. *Journal of Experimental Psychology: General*, 134(4), 565–584.
- Nosek, B. A., Banaji, M. R. & Greenwald, A. G. (2002). Math = Male, Me = Female, Therefore Math \neq Me. *Journal of Personality and Social Psychology*, 83(1), 44–59.
- Nosek, B. A., Greenwald, A. G. & Banaji, M. R. (2007). The Implicit Association Test at Age 7: A Methodological and Conceptual Review. In J. A. Bargh (Hrsg.), *Automatic processes in social thinking and behavior* (Kap. 6, S. 265–292). Psychology Press.
- Nosek, B. A., Greenwald, A. G. & Banaji, M. R. (2005). Understanding and Using the Implicit Association Test: II. Method variables and construct validity. *Personality and Social Psychology Bulletin*, 31, 166–180.
- Nosek, B. A. & Smyth, F. L. (2007). A Multitrait-Multimethod Validation of the Implicit Association Test: Implicit and Explicit Attitudes Are Related but Distinct Constructs. *Experimental Psychology*, 54(1), 14–29.
- Nosek, B. A., Smyth, F. L., Sriram, N., Lindner, N. M., Devos, T., Ayala, A., ... Greenwald, A. G. (2009). National differences in gender-science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academy of Science*, 106, 10593–10597.
- OECD – Organisation für wirtschaftliche Zusammenarbeit und Entwicklung. (2007a). *PISA 2006 – Schulleistungen im internationalen Vergleich: Naturwissenschaftliche Kompetenzen für die Welt von morgen*. Bertelsmann Verlag.
- OECD – Organisation für wirtschaftliche Zusammenarbeit und Entwicklung. (2007b). *PISA 2006 Volume 2: Data*. OECD. Zugriff unter <http://www.oecd.org/pisa/pisaproducts/pisa2006/39703566.pdf>
- Osborne, J., Simon, S. & Collins, S. (2003). Attitudes towards science: A review of the literature and its implications. *International Journal of Science Education*, 25(9), 1049–1079.
- Osgood, C. E. & Tannenbaum, P. (1955). The principle of congruity in the prediction of attitude change. *Psychological Review*, 62, 42–55.

- Østergaard, E., Dahlin, B. & Hugo, A. (2008). Doing phenomenology in science education: A research review. *Studies in Science Education*, 44(2), 93–121.
- Paschen, H. (Hrsg.). (2010). *Erziehungswissenschaftliche Zugänge zur Waldorfpädagogik*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Prenzel, M., Carstensen, C. H., Frey, A., Drechsel, B. & Rönnebeck, S. (2007). PISA 2006 – Eine Einführung in die Studie. In M. Prenzel, C. Artelt, J. Baumert, W. Blum, M. Hammann, E. Klieme & R. Pekrun (Hrsg.), *PISA 2006: Die Ergebnisse der dritten internationalen Vergleichsstudie* (Kap. 2, S. 31–59). Münster: Waxmann.
- Prenzel, M., Schütte, K. & Walter, O. (2007). Interesse an den Naturwissenschaften. In M. Prenzel, C. Artelt, J. Baumert, W. Blum, M. Hammann, E. Klieme & R. Pekrun (Hrsg.), *PISA 2006: Die Ergebnisse der dritten internationalen Vergleichsstudie* (Kap. 3.2, S. 107–124). Münster: Waxmann.
- Rausch, J. R., Maxwell, S. E. & Kelley, K. (2003). Analytic Methods for Questions Pertaining to a Randomized Pretest, Posttest, Follow-Up Design. *Journal of Clinical Child and Adolescent Psychology*, 32(3), 467–486.
- Rehfeldt, D. (2012). *Assoziationen zur Physik: Die Erprobung impliziter Testverfahren als Instrumente in der physikdidaktischen Forschung* (Masterarbeit, Humboldt-Universität zu Berlin, Institut für Physik).
- Rost, D. H. (2005). *Interpretation und Bewertung pädagogischer-psychologischer Studien: Eine Einführung*. Weinheim und Basel: Beltz Verlag.
- Rudolph, A., Schröder-Abé, M., Schütz, A., Gregg, A. P. & Sedikides, C. (2008). Through a Glass, Less Darkly? Reassessing Convergent and Discriminant Validity in Measures of Implicit Self-Esteem. *European Journal of Psychological Assessment*, 24(4), 273–281.
- Rutland, A., Cameron, L., Milne, A. & McGeorge, P. (2005). Social norms and self-presentation: Children's implicit and explicit intergroup attitudes. *Child Development*, 76, 451–466.
- Schiefele, U. (2009). Motivation. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (Kap. 7, S. 151–177). Heidelberg: Springer Medizin Verlag.
- Schnabel, K., Asendorpf, J. B. & Greenwald, A. G. (2007). Using Implicit Association Tests for the Assessment of Implicit Personality Self-Concept. In G. J. Boyle, G. Matthews & H. Saklofske (Hrsg.), *Handbook of Personality Theory and Testing* (Kap. 24, S. 508–528). Sage.
- Schnabel, K., Asendorpf, J. B. & Greenwald, A. G. (2008). Assessment of Individual Differences in Implicit Cognition: A Review of IAT Measures. *European Journal of Psychological Assessment*, 24, 210–217.
- Schnabel, K., Banse, R. & Asendorpf, J. B. (2006). Employing Automatic Approach and Avoidance Tendencies for the Assessment of Implicit Personality Self-Concept:

- The Implicit Association Procedure (IAP). *Experimental Psychology*, 53(1), 69–76.
- Schnell, R., Hill, P. B. & Esser, E. (1999). *Methoden der empirischen Sozialforschung* (8. Auflage). München, Wien: Oldenbourg Verlag.
- Schön, L. (1984). Bilder am großen Spiegel: Beispiele für eine Berücksichtigung affektiver Komponenten im Physikunterricht. *Naturwissenschaften im Unterricht: Physik/Chemie*, 12, 429–433.
- Schön, L. (1985). Physik als Erlebnis? *physica didactica*, 12, 27–33.
- Schön, L. (1994). Ein Blick in den Spiegel: Von der Wahrnehmung zur Physik. *Physik in der Schule*, 32(1), 2–5.
- Schön, L.-H. & Weber, T. (2001). Fachdidaktische Forschungen am Beispiel eines Curriculums zur Optik. In H. Bayrhuber, C. Finkbeiner, K. H. Spinner & H. A. Zwerger (Hrsg.), *Lehr- und Lernforschung in den Fachdidaktiken* (Bd. 3, S. 281–294). Forschungen zur Fachdidaktik. Innsbruck: Studienverlag.
- Senatsverwaltung für Bildung, Jugend und Sport Berlin. (2006). Rahmenlehrplan für die Sekundarstufe I: Physik.
- Shavelson, R. J., Hubner, J. J. & Stanton, G. C. (1976). Self-Concept: Validation of Construct Interpretations. *Review of Educational Research*, 46(3), 407–441.
- Sommer, W. (2005). *Zur phänomenologischen Beschreibung der Beugung im Konzept optischer Wege: Entwicklung und Erprobung einer Unterrichtsreihe für die gymnasiale Oberstufe* (L.-H. Schön & J. Grebe-Ellis, Hrsg.). Phänomenologie in der Naturwissenschaft. Berlin: Logos Verlag.
- Sommer, W., Grebe-Ellis, J. & Vogt, J. (2004). Zur Beugung: Von einfachen Freihandversuchen über die Laue-Kegel zum reziproken Gitter und zur Ewald-Kugel. *Physik und Didaktik in Schule und Hochschule*, 3(2), 67–86.
- Spreckelsen, K. (1992). Ordnen und Verstehen im physikalischen Bereich des Sachunterrichts. In R. Lauterbach, W. Köhnlein, K. Spreckelsen & E. Klewitz (Hrsg.), *Wege des Ordners: Vorträge des Arbeitstreffens zum naturwissenschaftlich-technischen Sachunterricht am 11. und 12. März 1991 in Berlin* (Bd. 2, S. 63–72). Probleme und Perspektiven des Sachunterrichts. Kiel: IPN.
- Sriram, N. & Greenwald, A. G. (2009). The Brief Implicit Association Test. *Experimental Psychology*, 56(4), 283–294.
- Steinle, F. (2004). Exploratives Experimentieren: Charles Dufay und die Entdeckung der zwei Elektrizitäten. *Physik Journal*, 3(6), 47–52.
- Steinle, F. (2005). *Explorative Experimente: Ampère, Faraday und die Ursprünge der Elektrodynamik* (M. Folkerts, Hrsg.). Boethius. Stuttgart: Franz Steiner Verlag.
- Theilmann, F. (2006a). Ein anderer Blick auf die prismatischen Farben. In H. Grötzebauch & V. Nordmeier (Hrsg.), *Didaktik der Physik: Kassel 2006*. Deutsche Physikalische Gesellschaft. DPG-Frühjahrstagung. Berlin: Lehmanns Media.

- Theilmann, F. (2006b). *Expeditionen in die Mechanik: Themen und Motive für erscheinungsorientierten Physikunterricht*. DRUCKtuell Druck- und Verlagsgesellschaft mbH.
- Theilmann, F. (2008). Der Gedanke im Kontext: Zur Charakterisierung einer erscheinungsorientierten Physik. *Der mathematische und naturwissenschaftliche Unterricht*, 61(4), 236–241.
- Thomson Reuters. (2009). Interview von ScienceWatch mit Anthony Greenwald. Zugriff März 2009, unter <http://archive.sciencewatch.com/inter/aut/2009/09-mar/09marGreen/>
- Wagenschein, M. (1976). *Die Pädagogische Dimension der Physik* (4. Auflage) (J. Muth & K. Schaller, Hrsg.). Grundthemen der Pädagogischen Praxis. Braunschweig: Georg Westermann Verlag.
- Wagenschein, M. (1989). *Erinnerungen für morgen: Eine pädagogische Autobiographie* (2. Auflage). Weinheim und Basel: Beltz Verlag.
- Wagenschein, M. (2009a). Also ist es wirklich wahr... [1978]. In H. C. Berg (Hrsg.), *Naturphänomene sehen und verstehen: Genetische Lehrgänge* [Das Wagenschein-Studienbuch](4. Auflage, Kap. 7, Bd. 4, S. 42–45). Lehrkunsstdidaktik. Bern: hep Verlag.
- Wagenschein, M. (2009b). Das Licht und die Dinge: Einführung in die Optik [1952]. In H. C. Berg (Hrsg.), *Naturphänomene sehen und verstehen: Genetische Lehrgänge* [Das Wagenschein-Studienbuch](4. Auflage, Kap. 26, Bd. 4, S. 116–117). Lehrkunsstdidaktik. Bern: hep Verlag.
- Wagenschein, M. (2009c). *Naturphänomene sehen und verstehen: Genetische Lehrgänge* [Das Wagenschein-Studienbuch] (4. Auflage) (H. C. Berg, Hrsg.). Lehrkunsstdidaktik. Bern: hep Verlag.
- Wagenschein, M. (2009d). Rettet die Phänomene! Der Vorrang des Unmittelbaren [1976]. In H. C. Berg (Hrsg.), *Naturphänomene sehen und verstehen: Genetische Lehrgänge* [Das Wagenschein-Studienbuch](4. Auflage, Kap. 20, Bd. 4, S. 96–108). Lehrkunsstdidaktik. Bern: hep Verlag.
- Wagenschein, M. (2010). *Verstehen lehren: Genetisch – Sokratisch – Exemplarisch* (5. Auflage). Beltz Taschenbuch. Weinheim, Basel: Beltz Verlag.
- Walpuski, M., Kauertz, A., Kampa, N., Fischer, H. E., Mayer, J., Sumfleth, E. & Wellnitz, N. (2010). ESNaS – Evaluation der Standards für die Naturwissenschaften in der Sekundarstufe I. In A. Gehrmann, U. Hericks & M. Lüders (Hrsg.), *Bildungsstandards und Kompetenzmodelle: Beiträge zu einer aktuellen Diskussion über Schule, Lehrerbildung und Unterricht* (Kap. 12, S. 171–184). Verlag Julius Klinkhardt.
- Watson, J., Mcewen, A. & Dawson, S. (1994). Sixth Form A Level Students' Perceptions of the Difficulty, Intellectual Freedom, Social Benefit and Interest of Science

- and Arts Subjects. *Research in Science & Technological Education*, 12(1), 43–52. eprint: <http://www.tandfonline.com/doi/pdf/10.1080/0263514940120106>
- Weber, T. (2003). *Kumulatives Lernen im Physikunterricht: Eine vergleichende Untersuchung in Unterrichtsgängen zur geometrische Optik* (H. Niedderer & H. Fischler, Hrsg.). Studien zum Physiklernen. Berlin: Logos Verlag.
- Weber, T. & Schön, L.-H. (2000). Spiegelwelt statt Reflexionsgesetz: Vorschläge zum Anfangsunterricht über Optik. *NiU*, 11(60), 30–36.
- Werner, J. (2000). *Vom Licht zum Atom: Ein Unterrichtskonzept zur Quantenphysik unter Nutzung des Zeigermodells* (H. Niedderer & H. Fischler, Hrsg.). Studien zum Physiklernen. Berlin: Logos Verlag.
- Westphal, N., Schön, L.-H. & Grebe-Ellis, J. (2011). Merkmale phänomenbasierten Physikunterrichts. In H. Grötzebauch & V. Nordmeier (Hrsg.), *PhyDid B - Didaktik der Physik: Münster 2011*. Deutsche Physikalische Gesellschaft. DPG-Frühjahrstagung. Berlin. Zugriff unter <http://www.phydid.de/index.php/phydid-b>
- Widodo, A. & Duit, R. (2004). Konstruktivistische Sichtweisen vom Lehren und Lernen und die Praxis des Physikunterrichts. *Zeitschrift für Didaktik der Naturwissenschaften*, 10, 233–255.
- Wigboldus, D. H. J., Holland, R. W. & van Knippenberg, A. (2006). *Single target implicit associations*. Unveröffentlichtes Manuskript.
- Winsteps 1.0.0 [Computer Software]. (2006). Beaverton, Oregon: Winsteps.com.
- Wright, N. A. & Meade, A. W. (2012). An exploration of cognitive ability contamination in the Implicit Association Test methodology. *Computers in Human Behavior*, 28, 393–399.
- Zwick, M. M. & Renn, O. (2000). *Die Attraktivität von technischen und ingenieurwissenschaftlichen Fächern bei der Studien- und Berufswahl junger Frauen und Männer*. Akademie für Technikfolgenabschätzung in Baden-Württemberg. Zugriff unter <http://d-nb.info/1026442850/34>